

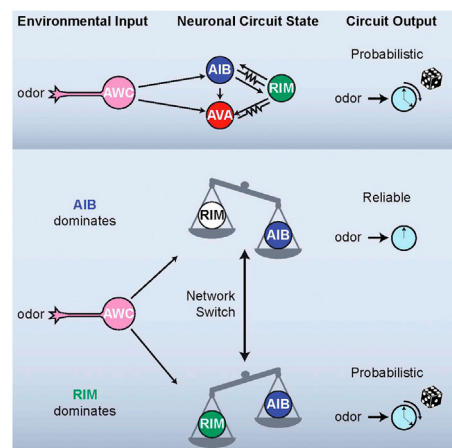


# Cell

Volume 161  
Number 2

April 9, 2015

[www.cell.com](http://www.cell.com)



## Circuit Stochasticity and Behavior

PAGE 215

Although animal behavioral responses to a stimulus are predictable on average, individual responses are highly variable. Gordus et al. study neuronal activity in response to an attractive odor in *C. elegans* and find that the stochastic relative activity states of interneurons downstream of sensory neurons determine the probability of the odor causing a motor output at a given time.

## Cancer in a Clamshell

PAGE 255

Massive loss of marine bivalve animals is happening throughout the world. Metzger et al. report that this is caused by a fatal form of cancer spreading through horizontal clonal transmission of tumor cells that likely arose from a single clam.

## A Gut Sense about Serotonin

PAGE 264

The gastrointestinal tract contains much of the body's serotonin, yet the control of its synthesis and breakdown is poorly understood. Yano et al. reveal that metabolites generated by spore-forming microbes in the gut microbiome promote host serotonin biosynthesis. Disrupting this communication impacts intestinal motility and hemostasis.

## All Together Now!

PAGE 277

Coordinated organ behavior is crucial for an effective response to environmental stimuli. Chen et al. demonstrate collective responses of hair follicles that allow an all-or-none decision to regenerate after plucking through a combination of molecular cues and immune cell recruitment.

## Unmasking Long Noncoding RNA-Protein Complexes

PAGE 404

The long noncoding RNA Xist is the master regulator of mammalian dosage compensation. Chu et al. develop a method to identify the composition of long noncoding RNA-protein complexes in vivo, providing insights into its temporal assembly and domain architecture.

## All about That Shape

PAGE 307

Although protein-DNA binding specificity is mediated by hydrogen bonds and hydrophobic contacts, these modes of recognition are not sufficient to explain specificity, particularly for factors with highly similar DNA binding domains. Abe et al. show that discrete DNA structural features play an independent and direct role in binding specificity by Hox proteins and that knowledge facilitates the de novo prediction of DNA binding specificities.

## A Dirty Sponge

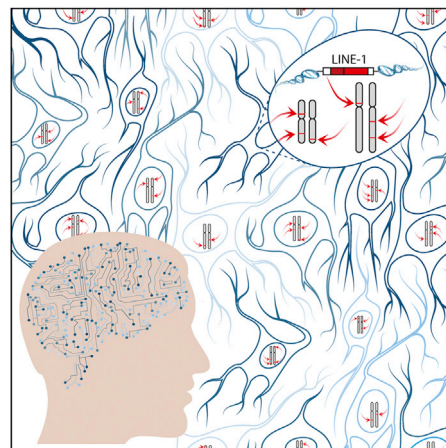
PAGE 319

Karreth et al. report that the BRAF pseudogene *BRAFP1* regulates the activity of its parental gene by competing for miRNA binding. Transgenic expression induces aggressive lymphoma in mice, and copy number analyses of human cancers further suggest an oncogenic function for the pseudogene.

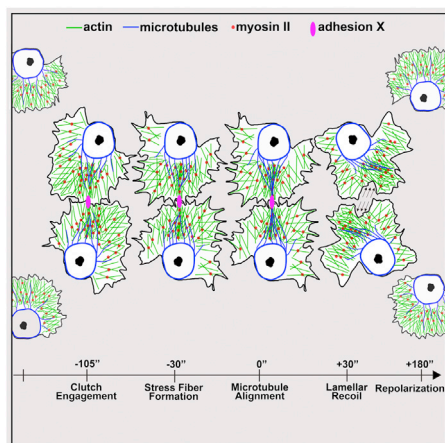
## Neuronal LINE Up

PAGE 228

Somatic genome mosaicism among neurons has the potential to impact brain function. Upton et al. show that LINE-1 retrotransposons mobilize extensively in hippocampal neurons, preferentially in hippocampally expressed loci. They are depleted from mature neurons when oriented in the most deleterious configuration to host genes, suggesting functional significance.







## Actin Fist Bump

PAGE 361

Cellular movement during development often originates when two cells collide and then repel each other, a process known as contact inhibition. Although it has been thought to be an uncoordinated reaction to contact, Davis et al. now show that it requires physical coupling of the flowing actin networks of the contacting cells. Actin acts as a mechanotransducer, allowing the cells to sense each other and coordinate their behaviors.

## A Little Trim Keeps Growth under Control

PAGE 333

Whereas ubiquitination targets many proteins for complete proteosomal degradation, modification of NF- $\kappa$ B p105 leads to selective proteolysis to generate p50. Kravtsova-Ivantsiv et al. identify the responsible E3 ligase and demonstrate that appropriate production of p50 is crucial for proper NF- $\kappa$ B signaling, as lowered levels of the ligase and p50 contribute to loss of growth control in tumorigenesis.

## Actin' Like a Pathogen

PAGE 348

Intracellular bacteria co-opt the host cell cytoskeleton to move in the cytoplasm and to disseminate. Benanti et al. find that pathogenic and non-pathogenic *Burkholderia* species use different strategies to drive actin-based motility. Pathogenic bacteria express a protein that mimics host Ena/VASP actin polymerases, which allows them organize actin filaments in a way that propels their movement and facilitates cell fusion during infection.

## Unified Theory of Migration

PAGE 374

Different cell types have intrinsically distinct migration patterns and responses to their environment. Maiuri et al. discover a simple rule that unifies the motile behavior of all cells: the straightness of movement or persistence is always an exponential function of their speed. Using this principle, they construct a physical model that predicts cell trajectories and reveals new insights about the molecular control of cell motility.

## Immunity and Genetics InterTWINed

PAGE 387

To understand how genetic variations may play role in the control of the homeostasis and disturbances of the immune system, Roederer et al. analyze 78,000 immune traits in 700 female twins and relate heritable traits with almost 300 single-nucleotide polymorphisms. The data link genetic control elements associated with normal immune traits to common autoimmune and infectious diseases, providing a shortcut to identifying potential mechanisms of immune-related diseases.

## Modeling Cancer with Stem Cells

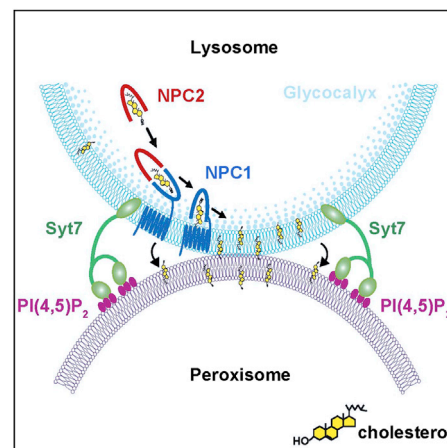
PAGE 240

Lee et al. establish a model of human familial cancer by deriving induced pluripotent stem cells from Li-Fraumeni syndrome patients. By modeling cancer with stem cells, they implicate the imprinted gene network in osteoblast differentiation defects and osteosarcoma development.

## Fumbling the Cholesterol Hand Off

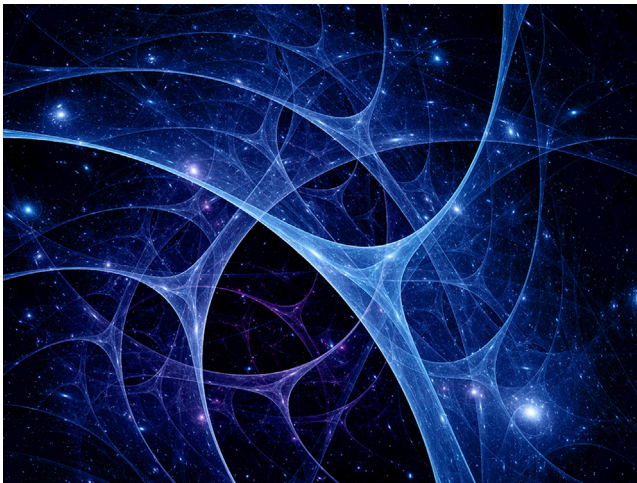
PAGE 291

How is cholesterol transported within cells? Chu et al. find that lysosomal LDL-cholesterol enhances contacts between lysosomes and peroxisomes to facilitate cholesterol transport and that cells from patients with peroxisomal disorders exhibit massive cholesterol accumulation.



# A Brief History of ALS

In a particularly poignant scene in the movie *The Theory of Everything*, a 21-year-old Stephen Hawking, full of excitement toward unlocking the secrets of the universe, learns of his debilitating diagnosis of amyotrophic lateral sclerosis (ALS). The year is 1963. He asks his doctor whether his brain will still be able to function in a landscape where the neurons controlling every aspect of his movement are rapidly dying. “Yes” says the doctor, unsure of what respite this piece of information could possibly give a young man facing a tragically short life ahead of him. Cut chase to 2015, and we are perhaps no closer to identifying the definitive cure for this particularly brutal but most common form of motor neuron disease. Only 10% of ALS cases are familial, and we do not yet know what the cause or causes can be for the majority of others. After diagnosis, patients typically do not survive any longer than 2–5 years. In this era of sequencing and big data, what we do have, however, is an increasingly useful treasure trove of information that can be mined to dissect out the complexity of this disease.



Charting neuronal function and connections over space and time. Image from iStock.com/alexovicsattila.

Mutations in SOD1 that lead to misfolding and intracellular aggregation of the enzyme are associated with 20% of familial cases of ALS (Robberecht and Philips, 2013). Although the best studied, the wealth of data on SOD1 has not brought the field much closer toward a viable therapeutic solution, largely due to the differences between disease progression in humans and animal models. In fact, there is a clear dearth of excellent animal models that are not based on overexpression of mutant proteins, and while iPSC-based methods have been helpful to model human disease, the need to study disease onset and progression in vivo has never been greater. Over the two decades, several other mutant genes have been linked to ALS, including those like SOD1 that aggregate and/or compromise cellular proteostasis, such as FUS,

ubiquilin2, and sequestosome (Robberecht and Philips, 2013).

Exome-sequencing studies, in particular, have proven to be invaluable in the identification of hereditary as well as de novo mutations associated with neurodevelopmental and neurodegenerative disorders. From the angle of familial ALS, Wu et al. identified profilin 1 in two independent familial cases of ALS (Wu et al., 2012) while Smith et al. sequenced 363 patient exomes to provide strong evidence for tubulin alpha 4a gene mutations (Smith et al., 2014). Both of these papers reveal a role for cytoskeletal regulators in driving disease. Aberrant RNA processing and the association of altered ribonucleoprotein homeostasis is another major emerging concept in ALS research. Mutations in the gene encoding the transcriptional and splicing regulator TDP-43 result in neuronal inclusions now regarded as a hallmark of the disease (Robberecht and Philips, 2013). Whether TDP-43 inclusions are causative or simply a characteristic of disease is not yet clear. In this regard, Johnson et al. identified mutations in the DNA and RNA-binding protein and TDP-43 interactor, MATR3, in several cases of familial ALS (Johnson et al., 2014).

Given, however, that the vast majority of diagnoses are of the sporadic form, such sequencing efforts should unravel equally, if not more valuable, data about what, if not possibly why, de novo mutations occur. Chesi and colleagues sequenced 47 ALS patients as well as their unaffected parents to identify a surprising enrichment for chromatin regulators, including *CREST* (Chesi et al., 2013). Most recently, Cirulli and group performed one of the largest ever ALS exome-sequencing studies by comparing 2,874 patients with over 6,000 controls (Cirulli et al., 2015). They identify with high confidence mutations in TANK-binding kinase 1 (TBK1), a kinase that controls key proteins in the autophagy pathway such as optineurin and p62, which themselves have been linked to ALS. Besides the models of altered proteostasis and ribostasis, Cirulli et al. propose that perhaps dysregulation of autophagy could be a central driving factor for disease progression.

As “mineable” sequencing data can be, they often raise many more questions than answers. The study from Cirulli et al. does not share significant overlap of top hits with other major sequencing efforts (Cirulli et al., 2015). What could be some of the reasons for these differences? Are familial and sporadic cases of ALS so very different? Is ALS a single disease or a spectrum of related diseases? Do patients need to be stratified based upon the severity of disease, age of diagnosis, sex, or any other parameters before such sequencing efforts are undertaken? Or do these individual studies put forth a broader conceptual point, that perhaps what really matters in understanding why motor neurons die is the specific pathway or the key biological process affected and not the specific genes? Researchers and clinicians are actively thinking about these very questions, and when one looks back at these sequencing studies in light of other ALS literature, an even more pertinent question comes to mind: how does one define ALS in a molecular context? Is it a disease of altered proteostasis or cellular quality control or is it really



more to do with RNA processing misregulation? Does it manifest because signaling events mediating interactions between key organelles at the heart of these biological processes, the ER, Golgi, and mitochondria are compromised? Or should one zoom out entirely at look at this at the cellular level to then look at the disease in the context of altered axonal function, neurite growth, or defects in excitation (Roselli and Caroni, 2015)? Indeed, a recent study has found that patient-derived motor neurons demonstrate increased membrane excitability irrespective of the mutation type (Wainger et al., 2014).

Collectively, these questions may seem daunting, even overwhelming, but in a way, they are also incredibly exciting in terms of developing several new frontiers for the exploration of ALS. Systems biology approaches in congruence with clinical and molecular studies will hopefully unravel and pair down key nodes that are most important for understanding as well as targeting the disease. Given that such approaches are being used to study other neurodegenerative diseases, it will also be interesting to study if there are shared commonalities between types of neurodegenerative disorders, such as the repeat expansions seen in the gene C9orf72 in ALS and frontotemporal dementia as well as Ataxin2 in spinocerebellar ataxia and ALS (Robberecht and Philips, 2013).

An enigmatic question that has plagued not just ALS but several other neurodegenerative disorders is the issue of selectivity. Assuming every cell in the body harbors the disease-causing or promoting mutation, why are motor neurons alone so vulnerable while other cell types, even neuronal classes spared? The concept of local insults adding fuel to a cell type already rather vulnerable, for reasons that are still not as clear, has been proposed to explain selective neuronal death (Roselli and Caroni, 2015). Perhaps localized neuroinflammation or damage to the local vasculature creates an environment where supporting cells fuel instead of healing damaged motor neurons. Cirulli et al. note that TBK1 also regulates the pro-inflammatory NF- $\kappa$ B pathway and that neuroinflammation may be an important mechanism to explore in the progression of disease (Cirulli et al., 2015).

A comprehensive picture of the disease is clearly imperative for rational drug design. Currently, riluzole is currently the main approved drug for the treatment of ALS, but its benefits are minimal. A number of compounds that showed promise in animal models failed clinical trials. A better understanding of what to target, as well as taking into account patient to patient variation, will be central to drug discovery and development. Having said this, for a relatively rare disease, ALS is thankfully not unheard of. The ALS Bucket Challenge, a social-media-fuelled phenomenon this past year alone helped to raise both awareness as well as over \$100 million toward the ALS Foundation. Patient care and management are additional challenges and will continue to be, especially if therapeutic interventions help to extend overall lifespan but without sufficiently retaining motor skills. Yet the right medical, technological, and social support can go a long way in preserving cognitive capacity as well as the overall quality of life of ALS patients, as exemplified by none other than Hawking himself, now 73, still solving mysteries of the universe.

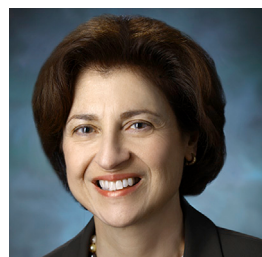
## REFERENCES

- Chesi, A., Staahl, B.T., Jovičić, A., Couthouis, J., Fasolino, M., Raphael, A.R., Yamazaki, T., Elias, L., Polak, M., Kelly, C., et al. (2013). *Nat. Neurosci.* **16**, 851–855.
- Cirulli, E.T., Lasseigne, B.N., Petrovski, S., Sapp, P.C., Dion, P.A., Leblond, C.S., Couthouis, J., Lu, Y.F., Wang, Q., Krueger, B.J., et al. (2015). *Science*. Published online February 19, 2015. <http://dx.doi.org/10.1126/science.aaa3650>.
- Johnson, J.O., Piro, E.P., Boehringer, A., Chia, R., Feit, H., Renton, A.E., Pliner, H.A., Abramzon, Y., Marangi, G., Winborn, B.J., et al; ITALSGEN Consortium (2014). *Nat. Neurosci.* **17**, 664–666.
- Robberecht, W., and Philips, T. (2013). *Nat. Rev. Neurosci.* **14**, 248–264.
- Roselli, F., and Caroni, P. (2015). *Neuron* **85**, 901–910.
- Smith, B.N., Ticozzi, N., Fallini, C., Gkazi, A.S., Topp, S., Kenna, K.P., Scotter, E.L., Kost, J., Keagle, P., Miller, J.W., et al; SLAGEN Consortium (2014). *Neuron* **84**, 324–331.
- Wainger, B.J., Kiskinis, E., Mellin, C., Wiskow, O., Han, S.S., Sandoe, J., Perez, N.P., Williams, L.A., Lee, S., Boulting, G., et al. (2014). *Cell Rep.* **7**, 1–11.
- Wu, C.H., Fallini, C., Ticozzi, N., Keagle, P.J., Sapp, P.C., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D.M., et al. (2012). *Nature* **488**, 499–503.

**Sri Devi Narasimhan**

# Immunotherapy: The Path to Win the War on Cancer?

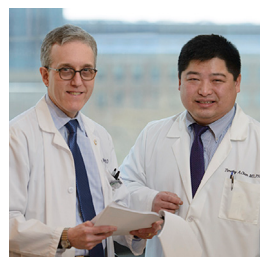
## On Breakthroughs and Evolution



**Suzanne L. Topalian**  
Johns Hopkins Kimmel Cancer Center

Drugs targeting immune checkpoint molecules such as CTLA-4, PD-1, and PD-L1 are being heralded as a breakthrough in oncology. However, “breakthrough” is a misnomer belying several decades of basic immunology advances and clinical trial and error leading up to this point. It was only after basic science uncovered the protean pathways restraining anti-tumor immunity that we could begin to unravel how to subvert them. The broad activity spectrum of PD-1 pathway blockers against multiple cancer types has validated this as a “common denominator” treatment approach and dispels the perception that “immunogenic” cancers are anomalies. We are now challenged to understand immune resistance mechanisms utilized by “non-responsive” tumor types (e.g., prostate cancer) as well as the 50% or more of individuals with “responsive” tumor types who do not benefit from these drugs. Identifying collateral pathways for co-targeting in combination treatment regimens requires an intellectual leap to consider unexpected intersections between the immune system and genetics, epigenetics, and metabolism. For instance, tumor mutational density, a surrogate indicator of neo-antigens available for immune recognition, correlates with the responsiveness of melanoma to anti-CTLA-4, and lung cancer to anti-PD-1. In the final analysis, teamwork with cross-fertilization of ideas across different scientific disciplines has driven the evolution to today’s “breakthroughs” and will meet tomorrow’s challenges.

## Central Dogma for Immunotherapy



**Jedd D. Wolchok and Timothy A. Chan**  
Memorial Sloan Kettering Cancer Center

In biology classes, we learned about the central dogma of molecular biology—DNA makes RNA and RNA makes protein. We’ve also learned about factors that regulate this central process, such as the influence of epigenetics, micro-RNAs, and mechanisms regulating post-transcriptional and translational control. Despite the fine intricacies, the central dogma of molecular biology remains intact—inherently elegant and graspable. A unifying concept for cancer immunology, on the other hand, has remained elusive until recently. We now have discovered the existence of molecular mechanisms of immune surveillance (thanks to Bob Schreiber) and that the number and quality of immune cells within the tumor microenvironment has significant prognostic impact in a variety of cancers. The quantity and quality of so-called “passenger” mutations in the tumor are also very important in determining the likelihood of success of immunologic checkpoint blockade with CTLA-4 or PD-1 pathway blocking antibodies. A putative dogma therefore is that mutations drive baseline immune reactivity and baseline immune reactivity is what determines the potential for benefit of immune potentiating therapies. As in molecular biology, there are likely to be modifiers, such as inhibitory cells populations, hostile microenvironments, and loss of antigen presenting capacity. Yet, a unifying concept will likely allow the field to further refine its approaches and specifically address the immunologic needs of individual patients.

## Not Just Another Hallmark



**Ira Mellman**  
Genentech

After an incubation period of nearly 100 years, cancer immunotherapy has emerged as a transformative approach to treat a wide variety of cancers. Although still early days, immunotherapy provides a degree of sustained clinical benefit rarely observed with more traditional cancer treatments. The excitement is, therefore, being largely driven by clinical results rather than by “breakthroughs” in the laboratory. There are nevertheless two daunting challenges. First, the field has progressed so rapidly in the clinic that our understanding of the underlying basic science and mechanisms of action are remarkably thin. Second, the tools we have to assess mechanism and correlates of treatment response (or lack thereof) remain rudimentary. Meeting these challenges is critical, since only a minority of patients as yet exhibit maximal benefit from immunotherapy. Importantly, clinical responses to agents such as anti-PD-L1/PD-1 are often clear and dramatic, thereby creating the opportunity to discover biomarkers and use them to understand inevitable patient to patient variations. Exploiting these correlates of clinical response will provide insights into basic cancer biology and inform immunotherapy combinations that can be expected to result in higher response rates and disease cures. Our task will be to backfill the science behind an exciting and validated therapeutic approach, ensuring that the field can look forward to a very exciting next decade both in the lab and in the clinic.



### New Trends in Cancer Vaccines



**Karolina Palucka and Jacques Banchereau**  
The Jackson Laboratory for Genomic Medicine

Clinical responses to checkpoint blockade are linked to the presence of T cell immunity to cancer-specific mutations. One way to increase the rate of clinical responses is to use vaccination to expand T cells specific for cancer mutations. Several phase III clinical trials testing different cancer vaccine candidates are currently ongoing. Exogenous vaccines utilize, for instance, dendritic cell-based and viral vectors-based approaches to boost the immune response in cancer patients. To be successful, these platforms will require applying high-throughput genomics to identify cancer-specific mutations and candidate peptide antigens in each patient in order to produce personalized vaccines. An alternative approach, endogenous vaccination, is based on exploiting the local release of antigens that happens upon standard cancer therapy (chemotherapy or radiotherapy) or oncolytic viral therapy. However, this strategy requires endogenous antigen presentation to be effective, in order to generate therapeutic T cell immunity. Dendritic cells are often skewed by tumors to generate pro-tumor immunity and thus reprogramming of their function in vivo is critical for the success of endogenous vaccination. Increasing the understanding of cancer genomics, the biology of antigen presentation and T cell biology will enable development of next-generation cancer vaccines which, combined with checkpoint blockade inhibitors, will pave the path to curative therapies for patients with cancer.

### Personalized Immunotherapy



**Steven A. Rosenberg**  
National Cancer Institute, NIH

Adoptive cell transfer (ACT) uses patient's lymphocytes to treat their autologous cancer. When tumor infiltrating lymphocytes are used for ACT, they can mediate complete, likely curative, regression in patients with metastatic melanoma. Lymphocytes genetically engineered to express anti-tumor receptors can treat patients with refractory lymphomas and leukemias. However, the majority of metastatic epithelial cancers are still resistant to immunotherapy. Recent approaches using deep exome sequencing along with high-throughput immunologic testing opened the door to treat these common types of cancer and to identify the rare somatic mutations that are immunogenic. ACT targeting these mutations is the ultimate "personalized" cancer treatment but is contrary to the mantra of many pharmaceutical companies who want "drugs in a vial" that can be mass produced and distributed. Although this approach has produced drugs that prolong the life of patients with solid metastatic cancers, curative treatments are rare and have had limited impact on overall death rates from cancer. A highly "personalized" immunotherapy for common cancers may require the development of a unique drug (autologous lymphocytes) for each patient. It will also need major changes and considerable flexibility on the part of industry. The effectiveness of treatment should trump simplicity of production and convenience of administration if progress is to be made in enabling patients with metastatic cancer to be cured and relish a normal lifespan.

### The New Immune Engineers



**K. Dane Wittrup**  
Koch Institute for Integrative Cancer Research, MIT

What vaccine best exploits the evolutionary weaknesses of a virus or a tumor's mutations? What is the intratumoral exposure history of intravenously injected antibodies? How does deregulated signaling tip the balance from healthy homeostasis to autoimmune disease? Can native T cell tropism overcome physical barriers to macromolecular drug delivery? How does our antibody repertoire respond to therapy or disease? How might innate and adaptive immunotherapies best be combined? How does lymphatic transport actively regulate adaptive immunity? Can injectable biomaterials program an effective anti-tumor immune response? A common thread through these varied and timely topics is the engagement of biological, chemical, and materials engineers at the forefront. At their disposal is an analytical toolkit honed to solve problems in the petrochemical and materials industries, which share the presence of complex reaction networks and convective and diffusive molecular transport. Powerful synthetic capabilities have also been crafted: binding proteins can be engineered with effectively arbitrary specificity and affinity, and multifunctional nanoparticles and gels have been designed to interact in highly specific fashions with cells and tissues. Fearless pursuit of knowledge and solutions across disciplinary boundaries characterizes this nascent discipline of immune engineering, synergizing with immunologists and clinicians to put immunotherapy into practice.

## Ronald J. Konopka (1947–2015)

Ron Konopka was found dead of an apparent heart attack in his Pasadena, CA home on February 14, 2015. Konopka was my close contemporary and began graduate school at Caltech in 1967. He published his thesis work along with his mentor Seymour Benzer in what is perhaps the single most influential paper in circadian rhythms (Konopka and Benzer, PNAS 68, 2112–2116). The field has spent much of the subsequent 45 years deciphering the meaning and validating (over and over again) the importance of this Rosetta stone. It began the modern era of circadian biology and is the cornerstone of my own circadian career. As if this were not enough, it is arguably *the* landmark paper in behavioral genetics writ large.

Benzer moved from Purdue to Caltech in the mid-60s and began this field; the physical move paralleled an intellectual move from prokaryotic genes to the underpinnings of behavior. He is properly credited with combining simple behavioral screens with the power of *Drosophila* genetics. The strategy could associate single mutations and the underlying genes with a behavioral phenotype. Although Benzer accumulated a coterie of talented students and post-docs to join him in this grand adventure, Konopka was the first. Moreover, he brought the circadian problem to Benzer rather than vice versa, and Ron designed as well as carried out the primary screen used to search for circadian mutants. The clock causes adult flies to eclose (emerge from the pupal case) at or shortly after dawn; this rhythmic emergence continues in constant darkness, with about 24 hr periodicity. The screen therefore searched for mutant flies that eclose in aberrant fashion and was remarkably successful. Ron found a short period mutant (about 20 hr), a long period mutant (about 30 hr) and an arrhythmic mutant.

Three striking features of the 1971 Konopka and Benzer paper led them to propose that the mutants were central to circadian rhythms. First, the three mutants affected not only the eclosion rhythm but also an independent circadian rhythm feature, locomotor activity, which also exhibited a short period, a long

period, or arrhythmicity. Second, genetic analysis indicated that all three mutations were alleles of a single gene, which they named *period*. The more expected result would have been three different genes each giving rise to the very different circadian phenotypes of fast, slow, or no rhythm; the finding of a single gene suggested that only a small number of gene products might be running the circadian clock. Third and most intriguingly, the results indicated that this single protein was of key importance for circadian timing, as it could mutate to a fast-running protein (short period) or a slow-running protein (long period) as well as being necessary for rhythmicity.

It took another 15 years for recombinant DNA and DNA sequencing to allow molecular characterization of the *period* gene and its protein, which verified some of these much earlier implications. For example, the short and long period alleles were determined to be missense mutations that altered the protein, whereas the arrhythmic mutation was a stop codon that prevented synthesis of the protein. Subsequent dynamic assays



Ronald J. Konopka

from many labs continue to this day and show that the short and long period alleles really do speed up and slow down the clock pace in ways that are being understood in considerable mechanistic detail. The *period* protein is also conserved in mammals. Although there are certainly some functional differences between the mammalian *period* proteins and the fly protein, one cannot overstate the extent to which the conclusions from Konopka and Benzer (1971)—drawn strictly from phenotypic and genetic studies—were prescient for the entire circadian field and all of its subsequent molecular sophistication.

Konopka did a post-doc at Stanford with the circadian biology pioneer Colin Pittendrigh and then was hired back at Caltech as an Assistant Professor in 1974.

Although publication requirements were much less onerous 40 years ago than today, Konopka was denied tenure based on his thin publication record from those assistant professor years. Nonetheless, important work from his lab was published at the end of his Caltech stay. Although these papers substantially added to the characterization of the *period* gene and its importance to circadian biology, they were deemed too late or insufficient to impact the tenure decision.

Konopka moved to Clarkson University in the early '80s. He had maintained a warm relationship with my long-time Brandeis collaborator Jeff Hall since their Caltech days and was important to our initial efforts to clone and identify the *period* gene. We were amateurs in the assays of locomotor rhythms, and Ron made sure that our first transgenic flies with wild-type *period* DNA constructs were indeed rhythmic. So we had truly rescued the arrhythmic behavior of the mutant host strain and had the gene in hand. Konopka continued to publish and was on track to receive tenure at Clarkson, but his promotion was apparently derailed by changed academic priorities at the university. He returned in 1990 to the small Pasadena house he had purchased while at Caltech.

Although Ron spent his last 25 years out of academic science, he began tutoring high school students in math and science after his return to Pasadena. According to his friend and former Benzer post-doc Larry Kauvar, "he was genuinely



fascinated by what makes science hard for some people and easy for others.” This long-standing commitment to teaching, along with a sardonic wit and broad interests, also contributed to his popularity as a Caltech professor, including by non-biologists. His hobbies included a first-rate butterfly collection as well as perhaps a thousand Grateful Dead concert tapes.

Few people know that Ron also played a seminal role in the beginnings of the Hereditary Disease Foundation. Milton Wexler, a psychoanalyst in Los Angeles, had begun to search for ways to attack Huntington’s disease (HD), an illness that affected his wife’s family. Wexler consulted with Benzer, who proposed in 1971 that Wexler hire his then 23-year-old graduate student Konopka. His task was to seek out talented people to attend a workshop and potentially pursue

research on HD. Konopka was so successful that Wexler hired him as the first Scientific Director of the organization that eventually became the Hereditary Disease Foundation. According to Alice Wexler, “Ron filled this post with his characteristic imagination and intelligence for several years. He played a wonderfully creative role in the history of the Hereditary Disease Foundation, and his legacy lives on to this day.”

Although Konopka participated only marginally in the molecular revolution that overtook behavioral genetics and fueled the remarkable progress of the circadian field since the mid-80s, his initial work was essential. The same is true for precious few researchers. Indeed, most scientists would fail the “deletion-test,” a term coined by Gerry Rubin to describe a scientist’s contributions by imagining what the field would be like had he/she

not existed. The same cannot be said of Konopka and his bold, revolutionary screen. That paper proved a very hard act to follow.

Sydney Brenner, and apparently J.D. Bernal before him, compared science to chess. They emphasized that the two games most worth playing are the opening game and the end game. Konopka and Benzer played the ultimate opening game. As Benzer died in 2007, Ron Konopka’s death closes this remarkable and singular chapter in the history of circadian rhythms, sadly the end of the beginning.

#### Michael Rosbash<sup>1,\*</sup>

<sup>1</sup>Department of Biology and Howard Hughes Medical Institute, Brandeis University, 415 South Street, Waltham, MA 02453, USA

\*Correspondence: [rosbash@brandeis.edu](mailto:rosbash@brandeis.edu)  
<http://dx.doi.org/10.1016/j.cell.2015.03.038>

# Pinning Down Randomness

William B. Kristan<sup>1,\*</sup>

<sup>1</sup>Section of Neurobiology, Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA

\*Correspondence: [wkristan@ucsd.edu](mailto:wkristan@ucsd.edu)

<http://dx.doi.org/10.1016/j.cell.2015.03.036>

**While some behavioral responses to a stimulus are invariant in animals, most are more likely to be variable or stochastic. In this issue, Gordus et al. illuminate a set of combinatorial neuronal activities that control the variability of a chemotactic behavior in response to an odor, providing a tractable system for understanding how stochastic circuit dynamics affect behavior.**

I was walking in a nearby desert some years ago on a beautiful Spring day, with wildflowers blooming all around and a breeze wafting from across a nearby grapefruit orchard in full bloom. The scent was transcendent, but every now and then, the breeze would slow down or change direction and the wonderful aroma would diminish or disappear. I found myself searching for the smell and soon realized that subconsciously I was employing an optimal strategy for finding patchy, unpredictably distributed targets. The strategy involves *random* back-and-forth searches and has been documented for predator-prey, pollinator-flower, and mating partner searches by such animals as reindeer, jackals, honey bees, seals, spider monkeys, microzooplankton, and Peruvian fishermen (Bertrand et al., 2007). Neuroscientists have noticed that learning often shows randomness both in the behavior and in electrophysiological recordings (Tumer and Brainard, 2007), but its source has been unknown. Until now. The research group headed by Cori Bargmann at the Rockefeller University (Gordus et al., 2015) has used the odor-searching behavior of the nematode worm *Caenorhabditis elegans* to track down the site and mechanism of this kind of behavioral variability.

Similar to my grapefruit odor seeking behavior, *C. elegans* (using only a few of its neurons) pursues attractive odors by moving forward as long as the intensity of the pleasant odor remains the same or increases but then changes or reverses directions if the gradient decreases (Pierce-Shimomura et al., 1999). The Bargmann lab folks studied the neurons that produce a reversal motion. The network that they studied consists

of just four pairs of neurons: chemosensory neurons (called AWC); motor output, or “reversal command” neurons (AVA); and two kinds of interneurons (AIB and RIM) in between (Figure 1).

This circuit suggests a mostly feedforward activation of the reversal motor neurons, and indeed, all eight neurons are activated whenever a reversal occurs. In addition, activating any one of the neurons optogenetically causes the number of reversals to increase. This result suggested to the Bargmann group that this circuit has a built-in variability generator, triggering random reversals of forward undulations.

To ensure that any variability in neuronal activity patterns was not due to diminished sensory perception but, rather, to circuit dynamics, the authors applied saturating concentrations of the odorant, which decreases the rate of reversals. Remarkably, the response to the odor did not remain constant for the whole duration of stimulus presentation (30–60 s) but instead flickered, jumping back and forth between no response (“off” state) and full response (“on” state). Even more remarkably, the whole circuit often flickered off and on at the same times. This flickering was correlated with the variability of the network’s output—the activation of the reversal command neurons. This correlation motivated the authors to determine which neurons were responsible for the variability.

Using a variety of techniques, the interneurons (AIB and RIM) and the reversal command neurons (AVA) were silenced, either individually (e.g., the pair of RIM interneurons) or in pairs (e.g., both AIB and RIM pairs). Strikingly, removing either type of interneuron made the behavior

more reliable: the response to the attractive odor more reliably inhibited the command neurons, and ablating both pairs of interneurons made the command neurons’ responses to the odor extremely reliable. These experiments were done on restrained worms using  $\text{Ca}^{2+}$  imaging to monitor neuronal activity, but the effect of eliminating one of the interneuron pairs—RIM—was confirmed to make the response to the odor more reliable in freely moving worms.

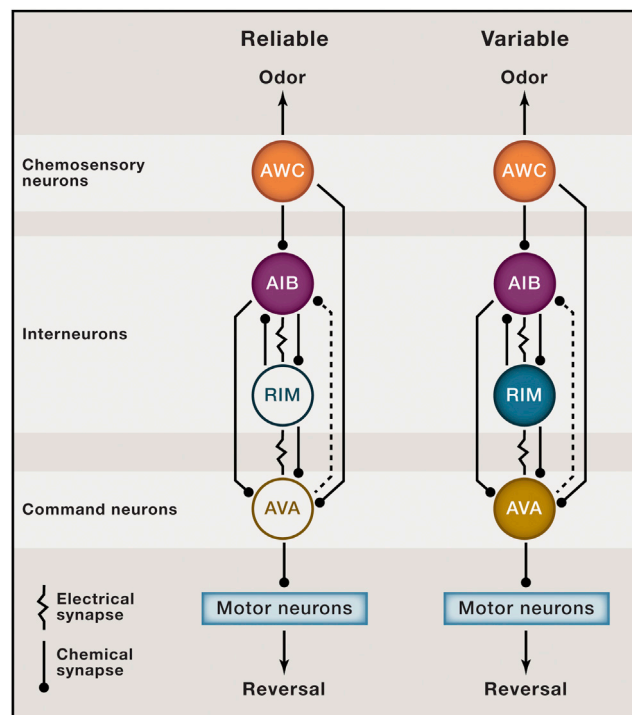
Although eliminating each of the interneurons has similar effects on the network’s output, their functions are not redundant. For instance, silencing the first interneuron in the chain (AIB) stabilizes the “off” state in the rest of the network, whereas silencing the second pair of interneurons (RIM) decreased the correlation of the flickers between the remaining interneurons (AIB) and the reversal command neurons (AVA). These results would not be predicted by the feedforward connections (AWC  $\Rightarrow$  AIB  $\Rightarrow$  RIM  $\Rightarrow$  AVA). Instead, these (and other) findings strongly suggest that the variability depends upon the feedback connections (from RIM to AIB and from AVA to AIB). Not surprisingly, eliminating chemical synaptic transmission in either pair of interneurons had the same effect on reducing variability as did silencing these neurons, indicating that it was the chemical synapses, not the electrical ones, that are responsible for the variability in the response.

Exactly how this network produces the variability is not clear. In part, that’s because the valence (excitatory or inhibitory) of the chemical synaptic connections is not entirely clear, especially in the feedback connections. For instance, the interneuron RIM releases



three different neurotransmitters (glutamate, acetylcholine, and tyramine), and the interneuron AIB releases only glutamate, but RIM (its primary target) expresses both excitatory and inhibitory receptors to this neurotransmitter. Doing electrophysiology in *C. elegans* is devilishly difficult, so working out the valence, strength, and temporal properties in these synaptic connections—as well as the inherent membrane properties—underlying the flickering awaits future studies.

It is interesting to consider why evolution might have inserted two layers of neurons in a circuit just to make that circuit's function less reliable. It could be entirely to implement the aforementioned optimal search pattern for food (Bartumeus et al., 2002), but it may have other functions too, such as providing a substrate for behavioral plasticity (Tumer and Brainard, 2007). In addition, the circuit shown in Figure 1 is embedded in more complex circuits in the worm's nervous system, such that the interneurons may be active in other behaviors. These neurons could act as traffic police, pointing neuronal activity toward different commanded behaviors. In line with this notion is the logical algorithm suggested by Gordus et al. for triggering reversals: the output state of the system (i.e., whether the reversal command neurons are activated or inactivated) depends upon the state of co-activation of the interneurons and the command neurons. The authors conclude that a reasonable hypothesis for explaining their data is that the flick-



**Figure 1. A Variability Generator Built into a Neuronal Circuit**

Four types of neurons (each type is paired) that influence reflexive reversal of locomotion, as well as their interconnections, are depicted. The names of the cells are based upon the location of their cell bodies in the nervous system, and the synaptic connections were determined from serial EM studies (White et al., 1986). The chemical synaptic contacts are likely to be mostly excitatory, acting on the neuron contacted by the filled circle. The electrical synapses allow electrical currents to pass both ways between neurons. The dashed connection was determined by experiments in this study, which may be indirect, through neurons not shown in this figure. The upward arrows from the AWC chemoreceptors indicate that removing an odor is the effective stimulus for eliciting reversals. The two states of the system shown are two of the most common three states of the circuit found immediately after odor is removed, thereby activating the chemosensory neuron (AWC). Color in a neuron means that it is in an activated state. In the “AIB-only” activated state, the odor essentially always elicits a behavioral reversal; hence, this state is reliable. When both pairs of interneurons (AIB and RIM) are active, the response becomes variable. (The third state—with AIB, RIM, and AVA all off—is also reliable.) These results indicate that the interneurons are the source of the variability in the response.

ering activity states of the neuronal network act like attractors, pulling the network into a particular combination of activity state that initiates and maintains the reversal behavior. Interestingly, the importance of attractor states in behavioral choice in both invertebrates (Briggman and Kristan, 2008) and vertebrates (Churchland et al., 2012; Mante et al.,

2013) has been recognized in recent years. Having this strategy present in such a simple nervous system as *C. elegans* is intriguing from an evolutionary point of view as well as for systems neurobiology, as it provides a comprehensible circuit for testing ideas about how attractor systems are put together and how they affect behavioral outcomes. Indeed, many other questions about combinatorial circuit dynamics and how and why they influence variable behavioral output in even broader contexts may now seem less daunting to tackle with the elegant system and conceptual framework provided by Gordus et al.

## REFERENCES

- Bartumeus, F., Catalan, J., Fulco, U.L., Lyra, M.L., and Viswanathan, G.M. (2002). *Phys. Rev. Lett.* 88, 097901.
- Bertrand, S., Bertrand, A., Guevara-Carrasco, R., and Gerlotto, F. (2007). *Ecol. Appl.* 17, 331–337.
- Briggman, K.L., and Kristan, W.B., Jr. (2008). *Annu. Rev. Neurosci.* 31, 271–294.
- Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). *Nature* 487, 51–56.
- Gordus, A., Pokala, N., Levy, S., Flavell, S.W., and Bargmann, C.I. (2015). *Cell* 161, this issue, 215–227.
- Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). *Nature* 503, 78–84.
- Pierce-Shimomura, J.T., Morse, T.M., and Lockery, S.R. (1999). *J. Neurosci.* 19, 9557–9569.
- Tumer, E.C., and Brainard, M.S. (2007). *Nature* 450, 1240–1244.
- White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). *Philos. Trans. R. Soc. Lond., B* 314, 1–340.

# The Clammy Grip of Parasitic Tumors

Robin A. Weiss<sup>1,\*</sup> and Ariberto Fassati<sup>1</sup>

<sup>1</sup>Wohl Virion Centre, Division of Infection and Immunity, University College London, Gower Street, London WC1E 6BT, UK

\*Correspondence: [r.weiss@ucl.ac.uk](mailto:r.weiss@ucl.ac.uk)

<http://dx.doi.org/10.1016/j.cell.2015.03.034>

**An epidemic of leukemia among bivalve molluscs is spreading along the Atlantic coast of North America, with a serious population decline of soft-shelled clams. In this issue of *Cell*, Metzger et al. use forensic DNA markers to demonstrate that the leukemia cells have a clonal origin and appear to be transmitted through sea water.**

Those of us who have enjoyed clam bakes on the beach at Cold Spring Harbor Laboratory may soon find that this mode of networking among scientists has become a thing of the past. The soft-shelled clam (*Mya arenaria*) is suffering a population dive owing to the spread of a lethal leukemia-like cell. Similar leukemias have been observed in other species of edible bivalve molluscs such as mussels, cockles, and oysters that are farmed along the Atlantic littoral of North America. The disease in soft-shelled clams now ranges from Chesapeake Bay to the Canadian province of Prince Edward Island 1,500 km to the northeast. Importation of clam stocks by shellfish farmers may have exacerbated its spread.

Steve Goff's laboratory at Columbia University, working with ecologists at Environment Canada, had previously identified a retrotransposon in the clams called *Steamer*, which is greatly amplified in the leukemic cells (Arriagada et al., 2014). Using genetic analysis of *Steamer* integration sites, mitochondrial single-nucleotide polymorphisms, and microsatellite variation, they now show that the leukemia has a monoclonal origin sharing common alleles that are different from their hosts (Metzger et al., 2015). In this respect, the leukemia in clams is similar to the Devil facial tumor disease (DFTD) of the marsupial carnivore, the Tasmanian devil, whose survival is endangered by rapidly spreading infection of a clonal neuro-endocrine cancer (Pearse and Swift, 2006; Murchison et al., 2012), and to canine transmissible venereal tumor (CTVT), which has a worldwide distribution in dogs (Murgia et al., 2006; Murchison et al., 2014).

DFTD is spread by biting, whereas CTVT, as its name implies, is sexually

transmitted. The mode of transmission of the clam leukemia is not yet firmly established, but it is likely that these filter feeders take up tumor cells from the seawater through their siphons (Figure 1) and that the cells then parasitize the new host. When the leukemic clone of clams first emerged is unknown, but since the disease was noted in the late 1970s, it must be at least 40 years old. DFTD was first recorded in Tasmanian devils in 1996 (Pearse and Swift 2006), whereas the venereal tumor in dogs is estimated to date from an ancient breed like the husky some 11,000 years ago (Murchison et al., 2014). Of course, even a 10,000 year period is a snapshot on the evolutionary timescale of the hosts, and one can speculate how many other cases of tumor cells evolving into parasites may have occurred in species that may now be extinct.

In addition to the three naturally occurring transmissible tumors of clams, dogs, and devils, there are several case histories of human malignancy arising from occult tumor cells in donor organ or tissue transplants that then emerge in the immunosuppressed transplant recipient (Murgia et al., 2006; Siddle and Kaufman, 2015). There are also cases of transplacental tumor transmission from mother to child and between twins in utero. Among inbred strains of laboratory rodents, there would be an opportunity for tumor cells to spread from one individual to another without crossing a major histocompatibility (MHC) barrier. However, only one example has been documented—that of leukemia in a colony of Syrian hamsters at NIH 50 years ago; the leukemic cells could even be transmitted by mosquitoes (Banfield et al., 1965), presumably by passive transfer on the mouth

parts rather than by undergoing a replication cycle in the insect host.

How do transmissible tumors manage to overcome histocompatibility barriers? It appears that there are a variety of mechanisms, including downregulation of class I and class II MHC genes and secretion of immunosuppressive cytokines (Belov, 2012; Siddle and Kaufman, 2015). In CTVT, there is a fine balance between progressive disease without an anti-tumor immune response and regression when allograft rejection kicks in. In DFTD, the tumor is relentlessly progressive. Both CTVT and DFTD have close to homozygous MHC class I alleles, and their initial emergence may have been facilitated by a relatively inbred host population with limited MHC diversity. Invertebrates like clams do not have as sophisticated a tissue recognition system as the MHC of higher vertebrates, yet certain cell surface molecules help to distinguish between self and non-self. However, non-malignant somatic cell invasion and even germ cell parasitism has been documented in marine invertebrates such as colonial tunicates (Rinkevich, 2011). Metzger et al. (2015) suggest that the lack of an MHC system may make molluscs more susceptible than vertebrates to transmissible tumors. As CTVT is the only known transmissible cancer that can regress, understanding what triggers rejection may be key to inducing regression of other transmissible tumors and perhaps non-transmissible cancers too.

As Metzger et al. (2015) document in their Introduction, leukemia occurs not only in soft-shelled clams, but also in other bivalve molluscs in the same region of North America. This observation raises the question of whether the



**Figure 1. A Bowl of Soft-Shelled Clams, *Mya arenaria*, with Extended Siphons**  
Courtesy of the Department for Natural Resources, State of Maryland.

tumors in other species represent cross-species infections from *Mya arenaria* or whether each host species has evolved its own transmissible tumor. CTVT is known to be readily transplantable experimentally to other canid species and even to foxes (Murgia et al., 2006; Belov, 2012), and it will be easy to determine the species of origin of transmissible leukemias in other bivalves. If each tumor is species specific, what environmental factors may have facili-

tated their independent emergence around the same time and place?

How might transmissible tumors be contained to reduce the threat to their host species? CTVT in dogs has spread to all five continents, but it is self-limiting and presents no danger to the host species as a whole. In India, the high incidence of CTVT in street dogs has been diminished in some cities by castration. To save the Tasmanian devil from extinction, a possible tumor vaccine is being

explored as well as containment of healthy animals on islands and on peninsulas with barrier fences to keep out affected devils (Belov, 2012). For the soft-shelled clam, it will be important to be vigilant to stop importation of clams from affected areas to currently unaffected ones like Florida.

In the meantime, it is safe for humans to eat clams, even raw oysters. Enjoy them while you can!

## REFERENCES

- Arriagada, G., Metzger, M.J., Muttray, A.F., Sherry, J., Reinisch, C., Street, C., Lipkin, W.I., and Goff, S.P. (2014). *Proc. Natl. Acad. Sci. USA* **111**, 14175–14180.
- Banfield, W.G., Woke, P.A., Mackay, C.M., and Cooper, H.L. (1965). *Science* **148**, 1239–1240.
- Belov, K. (2012). *BioEssays* **34**, 285–292.
- Metzger, M.J., Reinisch, C., Sherry, J., and Goff, S.P. (2015). *Cell* **161**, this issue, 255–263.
- Murchison, E.P., Schulz-Trieglaff, O.B., Ning, Z., Alexandrov, L.B., Bauer, M.J., Fu, B., Hims, M., Ding, Z., Ivakhno, S., Stewart, C., et al. (2012). *Cell* **148**, 780–791.
- Murchison, E.P., Wedge, D.C., Alexandrov, L.B., Fu, B., Martincorena, I., Ning, Z., Tubio, J.M., Werner, E.I., Allen, J., De Nardi, A.B., et al. (2014). *Science* **343**, 437–440.
- Murgia, C., Pritchard, J.K., Kim, S.Y., Fassati, A., and Weiss, R.A. (2006). *Cell* **126**, 477–487.
- Pearse, A.M., and Swift, K. (2006). *Nature* **439**, 549.
- Rinkevich, B. (2011). *Chimerism* **2**, 1–5.
- Siddle, H.V., and Kaufman, J. (2015). *Immunology* **144**, 11–20.



# Gut Microbiota: The Link to Your Second Brain

Vanessa Ridaura<sup>1,2</sup> and Yasmine Belkaid<sup>1,2,\*</sup>

<sup>1</sup>Program in Barrier Immunity and Repair

<sup>2</sup>Mucosal Immunology Section, Laboratory of Parasitic Diseases

National Institute of Allergy and Infectious Disease, NIH, Bethesda, MD 20892, USA

\*Correspondence: [ybelkaid@niaid.nih.gov](mailto:ybelkaid@niaid.nih.gov)

<http://dx.doi.org/10.1016/j.cell.2015.03.033>

**Serotonin is a highly ubiquitous signaling molecule that plays a role in the regulation of various physiological functions. Several lines of evidence, including the present work from Hsiao and colleagues, demonstrate that, in the gut, microbial-derived metabolites affect the production of serotonin that in turn impacts host physiological functions.**

The body's epithelial surfaces act as a scaffold to sustain diverse communities of commensals that include bacteria, archaea, fungi, protozoa, and virus. Although the notion that these microbial partners can promote human health is not a recent concept, the extent to which the microbiota controls all physiological systems has only recently begun to be appreciated. Of particular interest are the recent lines of investigation linking the microbiota to both the hormonal and nervous systems. In this context, a set of recent findings, including the work of Hsiao and colleagues, uncovers a role for the microbiota in controlling the production of a major neurotransmitter, serotonin.

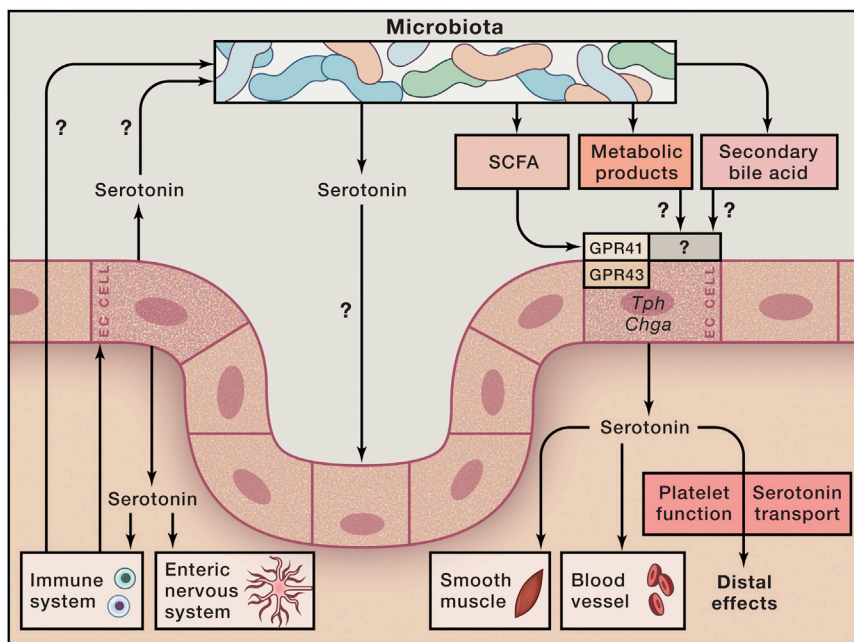
In 1967, Abrams and Bishop showed that animals devoid of live microbes (germ-free) had decreased gut motility compared to animals harboring a conventional mouse microbiota (Abrams and Bishop, 1967). Decreased serotonin levels in the absence of microbes have been proposed as a potential cause for this defect (Wikoff et al., 2009; Kashyap et al., 2013). Serotonin is a highly ubiquitous signaling molecule that plays a fundamental role in the regulation of various physiological functions via its action as both a neurotransmitter and a hormone. The vast majority of serotonin is produced in the gut by enterochromaffin cells, a specialized subset of cells strategically positioned to respond to chemical and mechanical stimuli (Figure 1). As a single molecule, serotonin has a remarkably wide range of effects on host physiology, ranging from the control of gut motility, secretory reflexes, platelet ag-

gregation, bone development, and cardiac function to the regulation of immune responses (Mawe and Hoffman, 2013). Germ-free mice display depressed levels of this neurotransmitter compared to conventionally raised animals (Wikoff et al., 2009). Colonization of germ-free animals with the gut microbiota from humans or mice can significantly accelerate gut transit time, an effect that can be partially blocked using a pharmacologic antagonist of a serotonin receptor (Kashyap et al., 2013).

In the present issue of *Cell*, the work of Yano et al. (2015), together with previous lines of evidence, proposes a model in which microbial-derived metabolites, such as short-chain fatty acids (i.e., butyrate and acetate) or secondary bile acids (specifically deoxycholate), directly act upon enterochromaffin cells (ECs), inducing transcription of the rate-limiting serotonin biosynthetic enzyme *thp1* (Yano et al., 2015; Fukumoto et al., 2003; Reigstad et al., 2014; Figure 1). A link between microbiota-derived metabolic products and enterochromaffin cell function has been previously reported by Reigstad et al., where they propose that short-chain fatty acids can modulate transcription of *Chga* gene, which encodes the neuroendocrine secretory protein chromogranin A, released together with serotonin (Figure 1). These results were corroborated by Yano et al. and correlated, in treated mice, with decreased transit time. Further, via increase in serotonin production, the presence of the microbiota was associated with increased platelet activation and aggregation, find-

ings that help to explain the improved coagulation at the site of injury in animals harboring a gut microbiota compared to germ-free mice. Of interest is the diversity of microbial-derived metabolites that can increase serotonin production, supporting the idea that the mechanisms controlling this brain-gut interaction are redundant and are unlikely to be affected by subtle microbial shifts. Because of the fundamental importance of serotonin in mediating central body functions, such redundancy may have been maintained as a means to sustain the production of this neurotransmitter in the face of constitutive microbial fluctuation. Nonetheless, not all microbes are equally efficient at promoting serotonin production. Yano et al. correlated the localized changes of serotonin with the presence of spore-forming bacteria, primarily from the *Clostridium* genus. Given the diversity of microbially derived metabolites, it is unlikely that a defined microbe will be associated with this effect. The mechanism by which the microbiota promotes serotonin production and, in particular, if and how enterochromaffin cells directly respond to microbiota-derived products and metabolites remain unclear. Furthermore, the systemic effects of this control on distal tissues and peripheral function are still unknown.

Of interest, and yet to be studied, is the effect that increased serotonin may have on the ecology and metabolism of the gut microbiota (Figure 1). Serotonin can act directly or indirectly on the immune system, which in turn can shape the microbiota composition and localization.



### Figure 1. Role of the Microbiota in Serotonin Production

The gut microbiota influences the number and function of enterochromaffin cells, thereby promoting the release of serotonin (5-HT). The microbiota promotes serotonin production via various metabolites, including short-chain fatty acid (SCFA). Serotonin can act on various physiological systems promoting gut motility, secretory reflexes, and platelet function. The action of serotonin can be local as well as distal (on bone development and cardiac function) via platelet-mediated transport. Serotonin can also influence the immune system, an effect that could feed back on enterochromaffin cells and the microbiota itself. Defined microbes have also been shown to produce serotonin, a pathway that may further link the microbiota and host serotonin levels.

The feedback loop associated with this control remains to be explored. Microbially induced changes in serotonin concentrations may regulate the host's immune response and may subsequently influence how the host deals with pathogens or commensals. Of interest, defined microbes themselves can produce serotonin (O'Mahony et al., 2015); this raises the question of whether serotonin can play a direct role in microbial metabolism or in the maintenance of the ecological niche of certain phylogenetic groups (i.e., spore-forming bacteria).

An important line of future investigation will be to explore how microbiota control of serotonin levels contributes to mucosal disorders, a question of particular importance because of the known role of serotonin in promoting immunity and inflammation. Of interest, in various models of mucosal inflammation or infections, as well as in Celiac disease patients, serotonin levels are significantly increased (Mawe and Hoffman, 2013). Further, both antagonists and agonists to the serotonin receptor have been used to clinically treat a variety of gastro-

intestinal disorders (Manocha and Khan, 2012). Based on the present findings, further understanding of the mechanism by which the microbiota regulates host's serotonin levels may be a first step toward developing pro- and/or prebiotic strategies to complement or potentially replace existing clinical treatments.

## ACKNOWLEDGMENTS

This work was supported by the Division of Intramural Research of the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health.

## REFERENCES

- Abrams, G.D., and Bishop, J.E. (1967). *Proc. Soc. Exp. Biol. Med.* 126, 301–304.
- Fukumoto, S., Tatewaki, M., Yamada, T., Fujimiya, M., Mantyh, C., Voss, M., Eubanks, S., Harris, M., Pappas, T.N., and Takahashi, T. (2003). *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 284, R1269–R1276.
- Kashyap, P.C., Marcobal, A., Ursell, L.K., Lar-  
auche, M., Duboc, H., Earle, K.A., Sonnenburg,  
E.D., Ferreyra, J.A., Higginbottom, S.K., Million,  
M., et al. (2013). *Gastroenterology* 144, 967–977.
- Manocha, M., and Khan, W.I. (2012). *Clin. Transl. Gastroenterol.* 3, e13.
- Mawe, G.M., and Hoffman, J.M. (2013). *Nature Rev. Gastroenterol. Hepatol.* 10, 473–486.
- O'Mahony, S.M., Clarke, G., Borre, Y.E., Dinan, T.G., and Cryan, J.F. (2015). *Behav. Brain Res.* 277, 32–48.
- Reigstad, C.S., Salmonson, C.E., Rainey, J.F., 3rd, Szurszewski, J.H., Linden, D.R., Sonnenburg, J.L., Farrugia, G., and Kashyap, P.C. (2014). *FASEB J.* Published online December 30, 2014. <http://dx.doi.org/10.1096/fj.14-259598>.
- Wikoff, W.R., Anfora, A.T., Liu, J., Schultz, P.G., Lesley, S.A., Peters, E.C., and Siuzdak, G. (2009). *Proc. Natl. Acad. Sci. USA* 106, 3698–3703.
- Yano, J.M., Yu, K., Donaldson, G.P., Shastri, G.G., Ann, P., Ma, L., Nagler, C.R., Ismagilov, R.F., Mazmanian, S.K., and Hsiao, E.Y. (2015). *Cell* 161, this issue, 264–276.

# A Collective Path toward Regeneration

Théo Maire<sup>1,2</sup> and Hyun Youk<sup>1,2,\*</sup>

<sup>1</sup>Department of Bionanoscience

<sup>2</sup>Kavli Institute of Nanoscience

Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, the Netherlands

\*Correspondence: [h.youk@tudelft.nl](mailto:h.youk@tudelft.nl)

<http://dx.doi.org/10.1016/j.cell.2015.03.039>

**How do cells collectively control an organ's behavior? By plucking various numbers of hairs from the mouse skin, Chen et al. show that hairs regenerate only when a sufficiently high density of them are plucked. Remarkably, a hair follicle can only regenerate in concert with other follicles, but not autonomously.**

A cell can modify its behavior in response to signaling molecules secreted by its neighboring cells. This cell, in turn, can secrete a signaling molecule that changes its neighboring cells' behaviors. In a population of cells, such a dynamic back and forth between many cells means that there are often no sharp boundaries between actions of cells that are autonomous and those that influence other cells. This blending of many individuals into one collective entity, often at a macroscopic scale, is a hallmark of multicellular systems like tissues, organs, and populations of bacteria. It also necessitates quantitative analyses to identify the cascades of events that yield collective behaviors of cells. In this issue of *Cell*, Chen et al. (2015) use mathematical models and experiments to reveal that a group of hair follicles can "count" how many hairs have been plucked from the skin and then regenerate the lost hairs if the density of plucked hairs is above a certain density (i.e., "threshold density") while not regenerating any hairs if the density of plucked hairs is below this threshold density (Figure 1A). The authors have thus uncovered a rare example of quorum sensing, which has mainly been studied in bacteria, at the level of a whole organ in a live animal.

Chen et al. (2015) studied hairs that grow on the mouse skin. A "hair unit" consists of a hair that protrudes from the skin and its follicle that lies beneath the skin (Figure 1A) (Jahoda and Christiano, 2011). The follicle contains the stem cells from which a new hair can grow. The authors counted and plucked individual hairs from mice. By varying the geometry of regions on the skin from which the hairs were plucked and the number of plucked hairs,

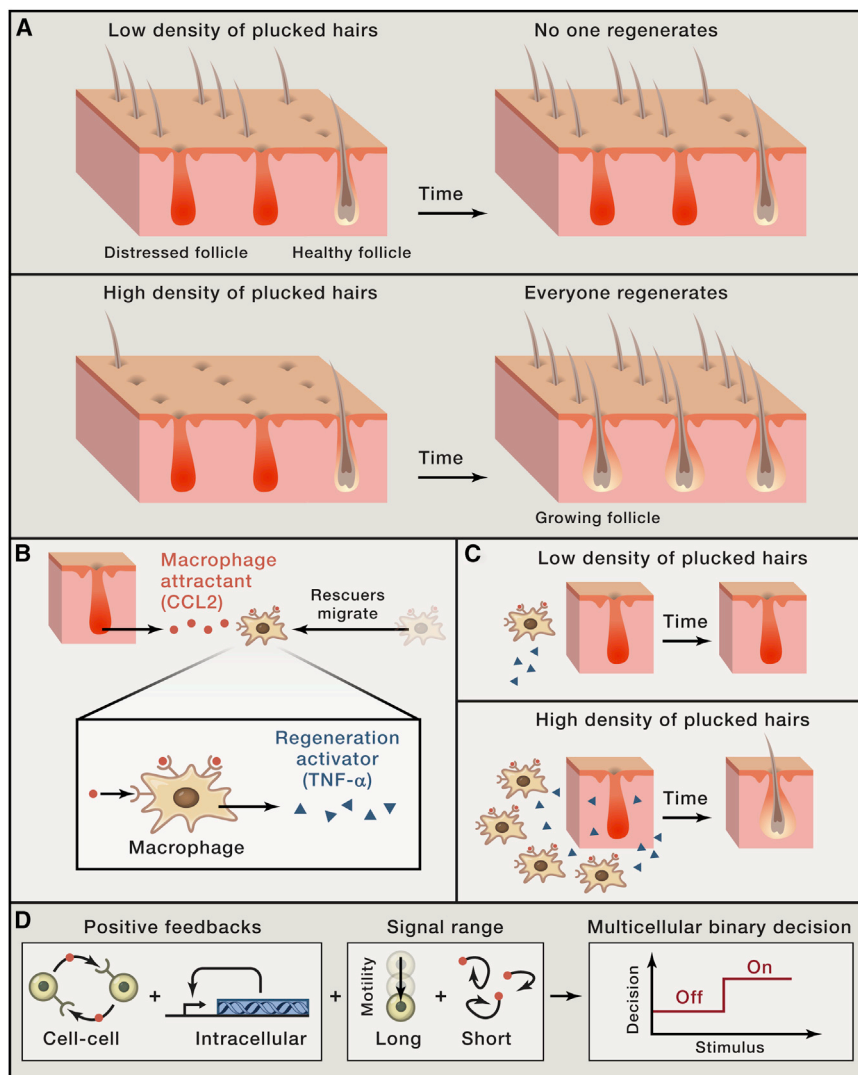
the authors discovered two scenarios. When the density of plucked hairs was below a certain threshold density, no follicles regenerated (Figure 1A). But when the density of plucked hairs was above the threshold density, both the follicles of the plucked hairs and of the surrounding intact hairs regenerated (Figure 1A). Thus, hair follicles are not autonomous. Instead, they collectively decide whether or not to regenerate both the lost and intact hair follicles. This ability of the follicles to measure the density of lost hairs is a form of quorum sensing (Ng and Bassler, 2009), in which a group of cells "measure" its population density and then together launch a collective action (i.e., regeneration of all hairs) only when the density is high enough. Using quorum sensing, the follicles can ignore harmless minor hair losses while using its resources to repair only harmful major hair losses.

Remarkably, Chen et al. (2015) discovered that a field of hair follicles spanning macroscopic distances (i.e., several mm) could quorum sense. Before degrading or being captured by a cell, a typical signaling molecule can travel no more than about 100  $\mu\text{m}$ . Thus, a "distressed" follicle (i.e., a follicle of a plucked hair) can potentially use a signaling molecule to tell its adjacent follicle, which is typically about 100  $\mu\text{m}$  away, of its hair loss. However, using a mathematical model, the authors deduced that follicles could not quorum sense across millimeter distances if they could only communicate with their immediate neighbors. In fact, they found that follicles must secrete a signal that traveled over a distance of 1 mm, at a higher speed than any molecule could achieve. Turning back to the

bench, the authors discovered that a distressed follicle recruited M1 macrophages to it by secreting the attractant chemokine CCL2 (Figure 1B). Immunostaining revealed that these motile macrophages first accumulated around the distressed follicles and then around the surrounding healthy ones. The macrophages secreted the signaling molecule  $\text{Tnf-}\alpha$  that stimulated the regeneration of both the healthy and distressed follicles through pathways that remain to be uncovered (Figure 1B). Although more work is required, the authors' data suggest that an appreciable accumulation of macrophages around the distressed follicles most likely occurs only when the density of distressed follicles is above the threshold density (Figure 1C). More importantly, the authors show that motile cells, along with signaling molecules, can transmit information over a macroscopic distance between immobile cells. An interesting question for the future is if such coupling between random diffusion of signaling molecules and directed motion of signaling cells may underlie collective behaviors of other organs.

An unresolved mystery is what determines the threshold density of hair loss. In microbial cells, the threshold that divides whether every cell or no cell responds to a signaling molecule is primarily determined by the binding affinity of the molecule to its receptor and a positive feedback regulation in a genetic circuit that controls the cells' response to the signaling molecule (Ng and Bassler, 2009; Pai et al., 2014; De Monte et al., 2007; Rotem et al., 2010). But, in multicellular systems, it is unclear how coordination of many different factors leads to a threshold and a binary





**Figure 1. A Hair Follicle Can Only Regenerate in Concert with Other Follicles, but Not by Itself**  
 (A) When the density of plucked hairs from the mouse skin is lower than a certain density (i.e., “threshold density”), no follicles regenerate. If the density of plucked hairs is higher than the threshold, all the follicles of plucked hairs (“distressed follicles”) and the follicles of surrounding intact hairs (“healthy follicles”) regenerate by entering growth (anagen) phase from a dormant (telogen) phase.  
 (B) A distressed hair follicle secretes the cytokine Ccl2. M1 macrophages sense Ccl2 and swim toward the distressed follicle.  
 (C) Macrophages secrete  $\text{Tnf-}\alpha$ .  $\text{Tnf-}\alpha$  activates regeneration of distressed and healthy follicles. Higher density of distressed follicles leads to a higher density of M1 macrophages recruited to the follicles.  
 (D) Main factors that a group of cells may use to collectively make a binary decision.

response (Figure 1D). In the case of the hair follicles, the short range of signaling mediated by diffusing molecules (e.g., Ccl2 and  $\text{Tnf-}\alpha$ ), the long range of signaling mediated by the motile macrophages, the spatial arrangements of the hair follicles, and the genetic circuits that control each cell’s secretion and response to the different signals must all fit together to set the threshold density of hair loss and a binary multicellular response (i.e., either

every follicle or no follicle regenerates) (Figure 1D). One possibility is that the macrophages make a binary decision (i.e., either move toward the distressed follicle or not) while the follicles are incapable of making any binary decisions. Another possibility is that a distressed follicle measures the density of macrophages surrounding it in such a way that it only regenerates when there is a sufficiently large density of macrophages.

Chen et al. (2015) and other recent studies (Hart et al., 2014; Sgro et al., 2015) motivate us to investigate how multiple cells, each with its own unique genetic circuit, can together achieve a collective function (e.g., a group of cells making a binary decision) that is analogous to certain behaviors of unicellular genetic circuits (e.g., a bistable genetic circuit). A key question is if there are other unicellular behaviors, which are governed by networks of genes that multicellular systems mimic with networks of communicating cells. A promising way to address this question is by building genetic circuits and cell-cell communications to reveal what sorts of multicellular behaviors can arise from them (Regot et al., 2011; Youk and Lim, 2014). An exciting outcome of this approach might be a realization that only a very small collection of genetic circuits and cell-cell interactions can yield a wide variety of collective behaviors of cells in nature.

#### ACKNOWLEDGMENTS

We thank S. Agrawal and B. Doganer for insightful suggestions.

#### REFERENCES

- Chen, C.-C., Wang, L., Plikus, M.V., Jiang, T.X., Murray, P.J., Ramos, R., Guerrero-Juarez, C.F., Hughes, M.W., Lee, O.K., Shi, S., et al. (2015). *Cell* 161, this issue, 277–290.
- De Monte, S., d’Ovidio, F., Danø, S., and Sørensen, P.G. (2007). *Proc. Natl. Acad. Sci. USA* 104, 18377–18381.
- Hart, Y., Reich-Zeliger, S., Antebi, Y.E., Zaretsky, I., Mayo, A.E., Alon, U., and Friedman, N. (2014). *Cell* 158, 1022–1032.
- Jahoda, C.A., and Christiano, A.M. (2011). *Cell* 146, 678–681.
- Ng, W.L., and Bassler, B.L. (2009). *Annu. Rev. Genet.* 43, 197–222.
- Pai, A., Srimani, J.K., Tanouchi, Y., and You, L. (2014). *ACS Synth. Biol.* 3, 220–227.
- Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellén, J., Peeters, T., Hohmann, S., de Nadal, E., Posas, F., and Solé, R. (2011). *Nature* 469, 207–211.
- Rotem, E., Loinger, A., Ronin, I., Levin-Reisman, I., Gabay, C., Shores, N., Biham, O., and Balaban, N.Q. (2010). *Proc. Natl. Acad. Sci. USA* 107, 12541–12546.
- Sgro, A.E., Schwab, D.J., Noorbakhsh, J., Mestler, T., Mehta, P., and Gregor, T. (2015). *Mol. Syst. Biol.* 11, 779.
- Youk, H., and Lim, W.A. (2014). *Science* 343, 1242782.

# Close Encounters of the Lysosome-Peroxisome Kind

Yui Jin,<sup>1,2</sup> Bethany S. Strunk,<sup>1,2</sup> and Lois S. Weisman<sup>1,\*</sup>

<sup>1</sup>Department of Cell and Developmental Biology, Life Sciences Institute, University of Michigan, Ann Arbor, MI, USA

<sup>2</sup>Co-first authors

\*Correspondence: [lweisman@umich.edu](mailto:lweisman@umich.edu)

<http://dx.doi.org/10.1016/j.cell.2015.03.046>

**Lysosomes provide a major source for cellular cholesterol; however, most of this cholesterol is trafficked to the plasma membrane via unknown mechanisms. Chu et al. identify an unexpected role for peroxisomes in the transport of cholesterol from the lysosome to the plasma membrane via a lysosome-peroxisome membrane contact site.**

Cholesterol is an essential determinant of membrane fluidity, permeability, and organization in animal cells. While the vast majority of cellular cholesterol (about 60%–80%) is localized at the plasma membrane (PM) (Maxfield and Wüstner, 2002), cholesterol originates from the ER via biosynthesis and the lysosome via import of exogenous cholesterol. These observations raise a fundamental question, how is cholesterol transported from the ER or lysosome to the PM? In this issue of *Cell*, Chu et al. (2015) discover that peroxisomes play a critical role in the transport of cholesterol from the lysosome to the PM and uncover an unexpected membrane contact site between the peroxisome and lysosome (Figure 1).

Exogenous cholesterol enters the cell as low density lipoproteins (LDL) via endocytosis of the LDL receptor. Upon delivery to the lysosome, LDL-derived cholesterol esters are de-esterified into free cholesterol then exported to other compartments including the PM (Maxfield and Wüstner, 2002). The physiological importance of cholesterol transport out of the lysosome is underscored by Niemann-Pick disease type C (NPC). NPC is a fatal, predominantly neurodegenerative disorder caused by mutations in NPC1 or NPC2, which results in cholesterol accumulation in the lysosome. NPC1 and NPC2 work together to transport free cholesterol out of the lumen to the limiting membrane of the lysosome (Du et al., 2011; Kwon et al., 2009; Vanier, 2015). The molecular mechanisms of subsequent steps, exit of cholesterol from the lysosomal membrane and delivery to the PM, were largely uncharacterized.

To identify proteins required for transport of LDL-derived cholesterol, Chu

et al. (2015) design an elegant screen that takes advantage of the antibiotic Amphotericin C, which permeabilizes the PM through association with exposed cholesterol, as well as U18666A, which enables them to stage the release of LDL-derived cholesterol from the lysosome. Using shRNA, they identify 341 candidate genes. Surprisingly, several candidates are related to peroxisomal function and biogenesis. Knockdown of these peroxisome related genes results in the accumulation of cholesterol in the lysosome lumen.

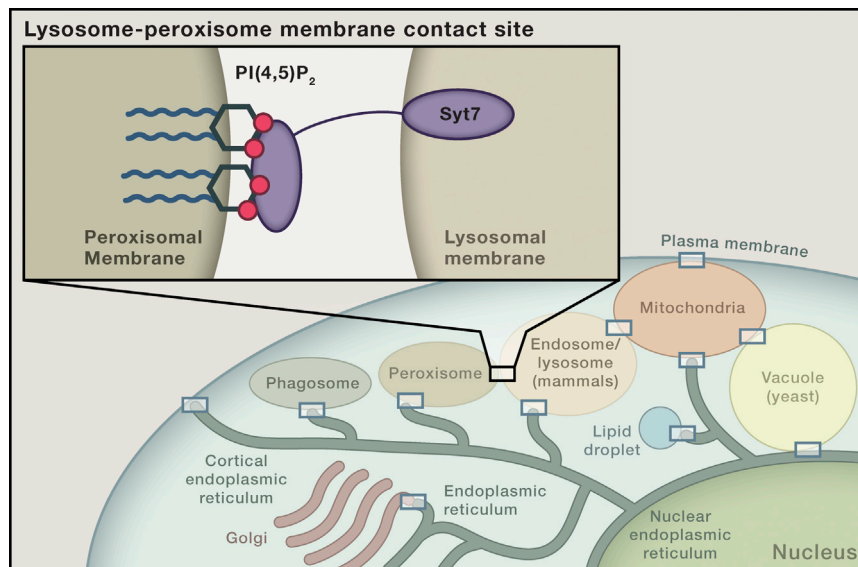
Analysis of cultured wild-type cells by both fluorescence microscopy and transmission electron microscopy reveals a previously unrecognized membrane contact site between lysosomes and peroxisomes. Further evidence for the lysosome-peroxisome contact site is provided by multiple in vitro studies demonstrating an interaction between these organelles. Chu et al. find that the lysosome-peroxisome contact site is bridged, at least in part, by the integral lysosomal membrane protein, synaptotagmin 7 (synt7), through binding to PI(4,5)P<sub>2</sub> on the peroxisomal membrane. The lysosome-peroxisome contact site is transient and cholesterol-dependent. Notably, efficient formation of the lysosome-peroxisome contact site also requires NPC1 suggesting that this contact is important for cholesterol exit from the lysosome.

Three of the peroxisomal genes identified in this study have been implicated in human diseases. X-linked adrenoleukodystrophy, Infantile Refsum disease, and Zellweger syndrome are caused by mutations in ABCD1, PEX1, and PEX26 (Aubourg and Wanders, 2013). Strikingly,

Chu et al. (2015) show that cells from patients suffering from each of these diseases accumulate cholesterol in lysosomes to a similar extent as those from NPC patients. The potential contribution of defects in cholesterol trafficking to symptoms of these diseases must now be considered.

The precise roles of the lysosome-peroxisome contact site in facilitating cholesterol transport out of the lysosome on its path to the PM remains to be determined. Chu et al. (2015) provide evidence from in vitro, as well as cell-based studies, that cholesterol is transferred from lysosomes to peroxisomes via the lysosome-peroxisome membrane contact site, raising the possibility that cholesterol may transit through the peroxisome on its way to the PM. It is also possible that the contact site facilitates transport of cholesterol out of the lysosome directly to a different organelle, such as the ER. In support of this possibility, knockdown of the oxysterol binding protein-related protein 5 (ORP5), which is localized to the ER, has been shown to result in accumulation of NPC1 dependent pools of cholesterol in the limiting membrane of the lysosome (Du et al., 2011). Orp5 may act in a parallel pathway, or downstream of Syt7 and ABCD1 in the transfer of cholesterol out of the lysosome, either to the ER or to the peroxisome.

Different pathways of cholesterol transport out of the lysosome may function in different cell types or under different conditions and perhaps only a subset of these pathways are directed to the PM. A comprehensive assessment of the molecular components of the lysosome-peroxisome contact, including the consideration



**Figure 1. The Lysosome-Peroxisome Contact Site Joins a Growing List of Inter-Organelle Contact Sites**

Known contact sites between the ER-mitochondria, ER-PM, ER-vacuole, ER-endosome/lysosome, ER-peroxisome, ER-Golgi, ER-phagosome, ER-lipid droplet, mitochondria-PM, and mitochondria-lysosome/vacuole are indicated.

of a possible three-way lysosome-peroxisome-ER junction, is necessary. This knowledge will be critical to understanding the peroxisome-dependent mechanisms of cholesterol transport out of the lysosome and for the development of disease therapies.

The lysosome-peroxisome contact site joins a growing list of inter-organelle contact sites. The known membrane contact sites currently include: ER-mitochondria, ER-PM, ER-lysosome/vacuole, ER-endosome, ER-Golgi, mitochondria-lysosome/vacuole, and mitochondria-PM (Prinz, 2014). Note that the lysosome-peroxisome contact site is now the third example of a critical contact between the lysosome and another organelle. Among the identified contact sites, some share partial functional redundancy. For example, the ER-mitochondrial encounter

structure (ERMES) and the vacuole and mitochondria patch (vCLAMP) are distinct membrane contact sites that connect the mitochondria to the ER and the yeast vacuole, respectively (Elbaz-Alon et al., 2014; Hönscher et al., 2014). Loss of either the ERMES or the vCLAMP has minimal phenotypic consequences, whereas simultaneous loss of both is lethal. Additional inter-organelle contacts are likely to be discovered.

In addition to specialized functions including calcium homeostasis and storage, intracellular signaling, organelle division, and lipid biosynthesis, membrane contact sites have repeatedly been shown to facilitate lipid transfer between membranes (Elbaz and Schuldiner, 2011; Prinz, 2014). It is tempting to speculate that inter-organelle contacts are the major routes of lipid transfer between

cellular compartments and fundamental to the accurate distribution of distinct lipid species throughout the cell. Currently, we are limited in our ability to observe the movements of lipids within cells because robust assays for tracking most lipids via microscopy do not yet exist. The development of methods facilitating such observations will be invaluable to advancement of the field.

#### ACKNOWLEDGMENTS

This work was supported by NIH grants R01-NS064015, R01-GM050403, and R37-GM062261 (to L.S.W.). B.S.S. was supported by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research.

#### REFERENCES

- Aubourg, P., and Wanders, R. (2013). *Handb. Clin. Neurol.* 113, 1593–1609.
- Chu, B.-B., Liao, Y.-C., Qi, W., Xie, C., Du, X., Wang, J., Yang, H., Miao, H.-H., Li, B.-L., and Song, B.-L. (2015). *Cell* 161, this issue, 291–306.
- Du, X., Kumar, J., Ferguson, C., Schulz, T.A., Ong, Y.S., Hong, W., Prinz, W.A., Parton, R.G., Brown, A.J., and Yang, H. (2011). *J. Cell Biol.* 192, 121–135.
- Elbaz, Y., and Schuldiner, M. (2011). *Trends Biochem. Sci.* 36, 616–623.
- Elbaz-Alon, Y., Rosenfeld-Gur, E., Shinder, V., Futerma, A.H., Geiger, T., and Schuldiner, M. (2014). *Dev. Cell* 30, 95–102.
- Hönscher, C., Mari, M., Auffarth, K., Bohnert, M., Griffith, J., Geerts, W., van der Laan, M., Cabrera, M., Reggiori, F., and Ungermann, C. (2014). *Dev. Cell* 30, 86–94.
- Kwon, H.J., Abi-Mosleh, L., Wang, M.L., Deisenhofer, J., Goldstein, J.L., Brown, M.S., and Infante, R.E. (2009). *Cell* 137, 1213–1224.
- Maxfield, F.R., and Wüstner, D. (2002). *J. Clin. Invest.* 110, 891–898.
- Prinz, W.A. (2014). *J. Cell Biol.* 205, 759–769.
- Vanier, M.T. (2015). *J. Inher. Metab. Dis.* 38, 187–199.



# Intracellular Bacteria Find the Right Motion

Edith Gouin,<sup>1,2,3</sup> Juan-Jose Quereda Torres,<sup>1,2,3</sup> and Pascale Cossart<sup>1,2,3,\*</sup>

<sup>1</sup>Unité des Interactions bactéries-Cellules, Institut Pasteur, 75015 Paris, France

<sup>2</sup>Inserm U604, 75015 Paris, France

<sup>3</sup>INRA USC2020, 75015 Paris, France

\*Correspondence: [pcossart@pasteur.fr](mailto:pcossart@pasteur.fr)

<http://dx.doi.org/10.1016/j.cell.2015.03.035>

**Benanti et al. report that *Burkholderia pseudomallei* and *Burkholderia mallei* bacteria express proteins that mimic Ena/Vasp family proteins to polymerize actin, thereby inducing actin-based motility. Thus, bacteria can use the various cellular actin polymerization mechanisms for intra- and inter-cellular dissemination.**

Intracellular bacterial pathogens such as *Listeria monocytogenes* and *Shigella flexneri* received a lot of attention in the early 1990s, when it was discovered that, after internalization by mammalian cells and escape from the endocytic vacuole, they actively recruit monomeric actin and polymerize it into filaments. This actin assembly process creates a force that then propels the bacteria through the cytosol. In this issue of *Cell*, Benanti et al. (2015) reveal that bacterial pathogens *Burkholderia pseudomallei* and *Burkholderia mallei* move inside of cells by mimicking one of the several cellular mechanisms that control actin polymerization.

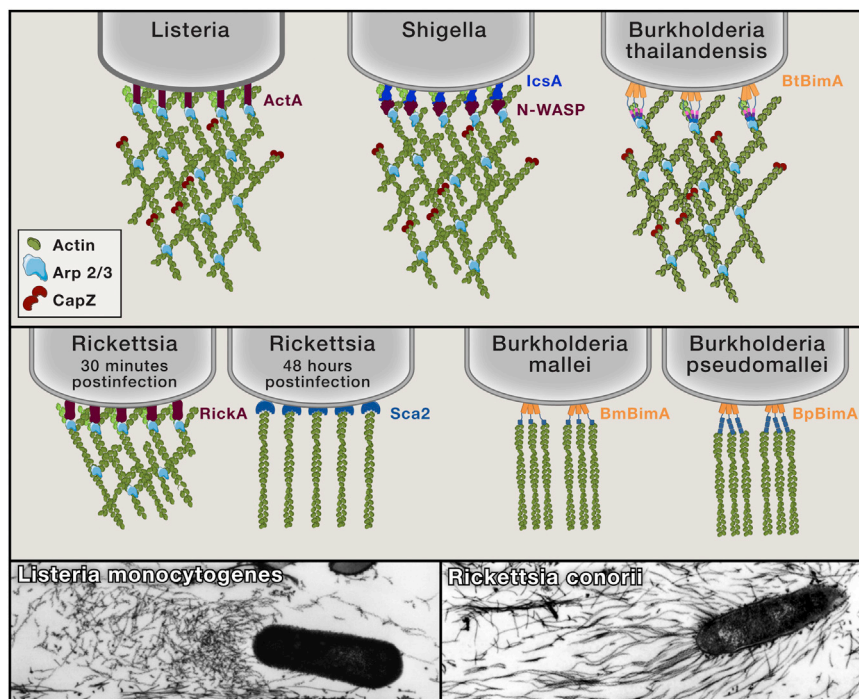
The study of bacterial motility led to the discovery of the role of Arp2/3 in actin polymerization (Welch et al., 1998). This seven-protein complex needs to be activated by WASP family proteins to generate a dense array of actin filaments. Polymerization of actin filaments takes place on preformed filaments at an angle of 70°, a characteristic feature that can be visualized in cells by electron microscopy or in vitro using fluorescent actin and Arp2/3. Other nucleators, such as formins or Ena/VASP proteins, as recently shown (Winkelman et al., 2014), generate linear long filaments. These proteins act processively on barbed ends in which they sit during the entire polymerization process. Interestingly, while *Listeria* and *Shigella* use Arp2/3, the intracellular pathogens *Rickettsia* early in infection use Arp2/3 and later on switch to a different mechanism, expressing proteins that mimic formins to promote actin-based motility (Welch et al., 1998; Suzuki et al., 1998; Gouin et al., 2004; Reed et al., 2014) (see below and Figure 1).

It is, however, the first time that bacterial mimics of Ena/VASP proteins are characterized (Benanti et al., 2015). The proteins encoded by the two *Burkholderia* species are called BpBimA and BmBimA and display three or one WH2 domains, respectively. WH2 domains bind to actin and are present in a number of actin-binding proteins, including Ena/VASP proteins. Two hallmarks of Ena/VASP proteins are their properties to oligomerize and to uncap actin filament barbed ends capped by capping proteins. Ena/VASP proteins can thus nucleate actin, elongate filaments at their barbed ends, and bundle them.

Previously, it was known that the non-pathogenic species *Burkholderia thailandensis* expressed a BimA protein called BtBimA and that Arp2/3 was necessary to induce actin-based motility. Here, the authors demonstrate that BtBimA has a VCA domain, known to bind and activate Arp2/3. In contrast, this domain is absent in BpBimA or BmBimA of the pathogenic species. A VCA domain is also present in WASP family proteins and in ActA of *Listeria* (Kocks et al., 1992). ActA is a bona fide mimic of WASP family proteins (Welch et al., 1998) (Figure 1). In the case of *Shigella*, the outer-membrane protein IcsA/VirG (Bernardini et al., 1989) recruits N-WASP (Suzuki et al., 1998), which in turn recruits and activates Arp2/3 (Figure 1). In *Rickettsia*, early during infection, the protein RickA, which displays a VCA domain, mimics WASP family proteins and recruits Arp2/3 (Gouin et al., 2004; Reed et al., 2014). Later, the surface protein Sca2 acts as a formin (Reed et al., 2014). As Sca2, BimA proteins belong to a family of bacterial proteins called “auto-

transporters” because they have the capacity to insert themselves in the outer membrane and then display their N-terminal parts on the outside of the bacterium. To this end, they trimerize via a coiled-coil domain that generates a pore through which transport does occur. In the case of BimA proteins, trimerization results in the exposure on the surface of nine WH2 domains for *B. pseudomallei* and three WH2 for *B. mallei*.

In their study, Benanti et al. perform a series of in vitro assays with purified proteins, mutants analysis, observations of comet tails, and plaque assays to analyze the dissemination of the bacteria from one infected cell to its neighbors. They find that BtBimA polymerizes actin if Arp2/3 is present but does not in its absence. In contrast, the two other BimAs are able to polymerize actin in the absence of Arp2/3. They also demonstrate that the two proteins remain associated with the barbed ends as the filaments grow. The affinities of BpBimA and BmBimA for the barbed ends are in a similar range to those of formin and Ena/VASP for barbed ends, supporting the idea that they are functionally related. In addition, the two BimAs can gather two filaments and elongate them as VASP does, as well as displace CapZ from the barbed ends. By creating mutations in the coiled-coil domain, the authors establish that trimerization is critical for actin nucleation, barbed end elongation, and anti-capping activity. They then compare the tails produced by *B. thailandensis*, *B. pseudomallei*, and *B. mallei* and find that they are curved for *B. thailandensis* and straight for the two other species. Importantly, the efficiency of movement (length



**Figure 1. Bacterial Actin-Based Motilities**

(Top) Schematic representation of the actin-based motilities of *Listeria*, *Shigella*, *Rickettsia* spp., and *Burkholderia* spp.

(Bottom) Electron micrographs of *L. monocytogenes* tails (left) and *R. conorii* tails (from Gouin et al., 1999 and our unpublished data).

of displacement) is higher in the case of *B. pseudomallei* and *B. mallei*, as these bacteria moved in a linear path in contrast to *B. thailandensis*, which moved in a curved path, with potential implications for the pathogenicity of these species.

In summary, this study unveils how bacteria can exploit different mechanisms offered by the host cell to polymerize actin. In fact, the results raise the possibility that other bacteria such as *Rickettsia*, although belonging to the same genus,

may also use different mechanisms to move inside of cells. Finally, it remains possible that BimA proteins play other roles in infection, as is the case of ActA, which covers the bacterial surface and as a Trojan horse protects *Listeria* from autophagy (Yoshikawa et al., 2009),

## REFERENCES

- Benanti, E.L., Nguyen, C.M., and Welch, M.D. (2015). *Cell* 161, this issue, 348–360.
- Bernardini, M.L., Mounier, J., d'Hauteville, H., Coquis-Rondon, M., and Sansonetti, P.J. (1989). *Proc. Natl. Acad. Sci. USA* 86, 3867–3871.
- Gouin, E., Gantelet, H., Egile, C., Lasa, I., Ohayon, H., Villiers, V., Gounon, P., Sansonetti, P.J., and Cossart, P. (1999). *J. Cell Sci.* 112, 1697–1708.
- Gouin, E., Egile, C., Dehoux, P., Villiers, V., Adams, J., Gertler, F., Li, R., and Cossart, P. (2004). *Nature* 427, 457–461.
- Kocks, C., Gouin, E., Tabouret, M., Berche, P., Ohayon, H., and Cossart, P. (1992). *Cell* 68, 521–531.
- Reed, S.C., Lamason, R.L., Risca, V.I., Abernathy, E., and Welch, M.D. (2014). *Curr. Biol.* 24, 98–103.
- Suzuki, T., Miki, H., Takenawa, T., and Sasakawa, C. (1998). *EMBO J.* 17, 2767–2776.
- Welch, M.D., Rosenblatt, J., Skoble, J., Portnoy, D.A., and Mitchison, T.J. (1998). *Science* 281, 105–108.
- Winkelman, J.D., Bilancia, C.G., Peifer, M., and Kovar, D.R. (2014). *Proc. Natl. Acad. Sci. USA* 111, 4121–4126.
- Yoshikawa, Y., Ogawa, M., Hain, T., Yoshida, M., Fukumatsu, M., Kim, M., Mimuro, H., Nakagawa, I., Yanagawa, T., Ishii, T., et al. (2009). *Nat. Cell Biol.* 11, 1233–1240.

# Immunoengineering: How Nanotechnology Can Enhance Cancer Immunotherapy

Michael S. Goldberg<sup>1,\*</sup>

<sup>1</sup>Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA

\*Correspondence: [michael\\_goldberg1@dfci.harvard.edu](mailto:michael_goldberg1@dfci.harvard.edu)

<http://dx.doi.org/10.1016/j.cell.2015.03.037>

Although cancer immunotherapy can lead to durable outcomes, the percentage of patients who respond to this disruptive approach remains modest to date. Encouragingly, nanotechnology can enhance the efficacy of immunostimulatory small molecules and biologics by altering their co-localization, biodistribution, and release kinetics.

## Awakening the Immune System

Although the research community has made great inroads into understanding the underlying etiology of cancer, our ability to confer durable responses to patients remains rather limited. The complexity of cancer aside, a major obstacle impeding our progress has been the widespread emphasis on cancer as a cell-autonomous disease. Few biologists would study gill physiology by removing a fish from water, yet we routinely interrogate cancer cells outside of their natural habitat, discounting the importance of the tumor microenvironment. In addition to stromal cells and extracellular matrix, immune cells greatly impact disease initiation, progression, and invasion.

Indeed, the type, density, and location of immune cells within tumors predict patient survival as well as, if not better than, traditional histopathological methods. This so-called “immune contexture”—most notably the presence of CD8<sup>+</sup>CD45RO<sup>+</sup> T cells and T<sub>H</sub>1 cells—is associated with a good prognosis across at least 20 different cancer types (Fridman et al., 2012). Accordingly, oncologists are eager to arouse exhausted immune cells, and clinical data confirm that stimulating a patient’s natural antitumor immune response can cure relapsed, refractory patients with difficult-to-treat cancers who have exhausted other treatment options (Topalian et al., 2011).

Challengingly, tumors can evade immune surveillance. Consequently, most immunotherapies, particularly those directed against solid tumors, have thus far benefited only a minority of patients. For this reason, facilitating antitumor immune cells to overcome the activation energy barrier presented by the immunosuppressive tumor microenvironment is an area of active investigation. Emerging preclinical and clinical data suggest that delivery of immunostimulatory molecules from nanoparticles and scaffolds can rouse the immune system with greater rigor than delivery of these same molecules in solution, leading to improved antitumor immunity and survival outcomes. Accordingly, biologists and engineers are working to improve our understanding of which cells and pathways should be perturbed to maximize efficacy and what tools are most appropriate to perturb them as desired.

## The Killer App for Nanomedicine?

Nanoparticles are synthetic particles (generally derived from polymers, lipids, or metals) with sizes on the nanometer scale,

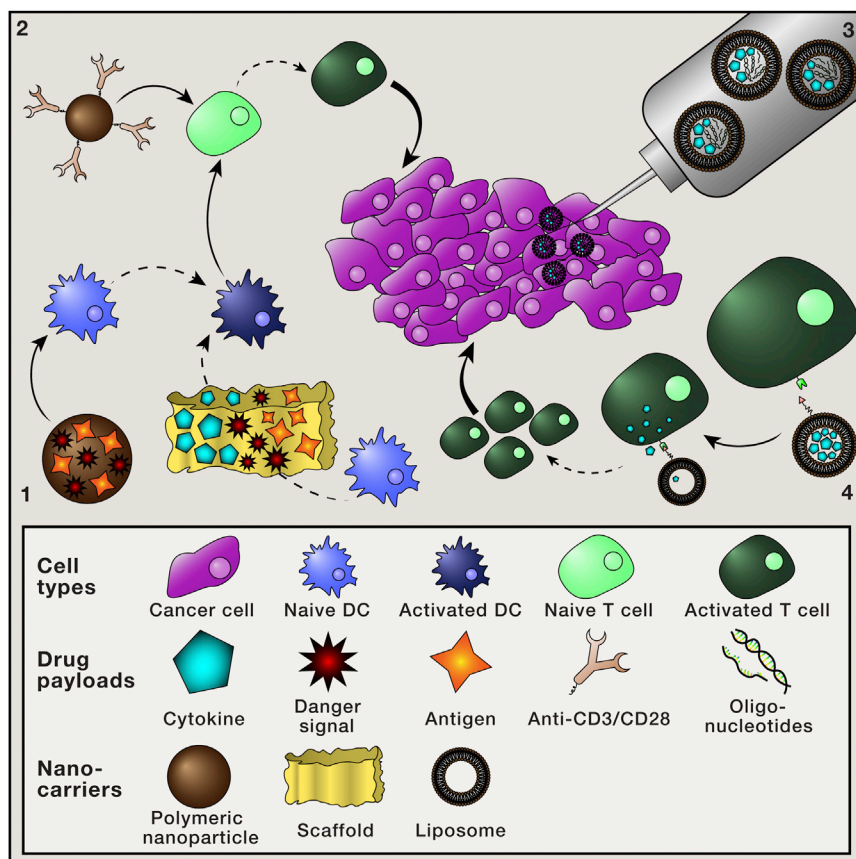
which confers properties that bridge bulk and molecular structures. Such nanoparticles can be loaded with therapeutic compounds to achieve concentrated local drug delivery with potential for sustained release when biodegradable carriers are used. Their high surface-area-to-volume ratio enables them to be coated with various ligands (e.g., antibodies or aptamers) that can facilitate interaction with cognate molecules, including receptors present on the surface of target cells. Although nanoparticles can improve the pharmacokinetic properties of their drug payloads (Chow and Ho, 2013), their ability to target cancer cells specifically and efficiently has proven somewhat elusive. Targeting nanoparticles to specific receptors on cancer cells augments cellular uptake but not tumor localization, which is governed by passive accumulation through leaky vasculature. In contrast, leukocytes can actively traffic to tumors along chemokine gradients, rendering these cells the ultimate “targeted” therapy.

Delivery of immunostimulatory drugs to antitumor immune cells may be a more efficient tactic to eradicate tumors than delivery of cytotoxic drugs to cancer cells (Figure 1). While the ability to concentrate nanoparticles within tumors upon systemic administration remains a challenge, immune cells proliferate extensively upon activation. As a consequence, unlike for cancer cells, successful payload delivery to even a small fraction of immune cells can achieve robust antitumor efficacy. Moreover, tumors are heterogeneous and ever evolving, so drugs that are designed to kill cancer cells directly by targeting cell-intrinsic pathways inherently select for resistant clones that lead to relapse. In contrast, immune cells can generate a coordinated and adaptive antitumor response with capacity for memory that is not achievable using any other therapeutic modality.

## Improving the Efficacy of Cancer Vaccines

Dendritic cells (DCs) are critical initiators of adaptive immune responses and are thus extremely relevant targets for anticancer nanomedicines. Co-administration of antigen and adjuvant as free drugs can result in delivery of antigen to some DCs and adjuvant to others. Delivery of antigen in the absence of adjuvant induces immunologic tolerance, thereby inhibiting robust antitumor responses. Co-encapsulation of antigen and adjuvant in a common particle enables co-delivery of both components to the same DC, leading to improved induction of antigen-specific





**Figure 1. Applications of Nanotechnology**

Clockwise from bottom left: nanocarriers can be used to (1) deliver cancer vaccine antigens and adjuvants to dendritic cells, (2) stimulate T cells directly as artificial antigen presenting cells, (3) concentrate immunostimulatory compounds in the immunosuppressive tumor microenvironment, and (4) deliver supportive drugs to T cells in the circulation. (Image credit to Mohammad H. Saleh.)

biology. First, owing to the fenestrated architectures of secondary lymphoid organs, nanoparticles naturally accumulate in these structures—particularly the spleen—which are populated by many DCs. Second, secondary lymphoid organs do not exhibit the physical barriers to entry that are characteristic of solid tumors, such as elevated interstitial pressure and impaired diffusion caused by unusually dense extracellular matrix.

Of note, the subset of DC targeted is critical to defining the induction and regulation of immune responses (Ueno et al., 2011). Plasmacytoid DCs can be converted from toleragenic to innate immunostimulatory upon uptake of Toll-like receptor (TLR) 7 and/or 9 agonists. To achieve adaptive responses, distinct subsets of classical DCs can be targeted by nanoparticles derived from poly(lactic-co-glycolic acid) (PLGA)—a biodegradable, FDA-approved polymer—to which antibodies are coupled. The C-type lectin receptor on the surface of the DC that is targeted by the antibody defines the type of immune response produced. For example, targeting of DC-SIGN, DEC-205, DNGR-1, and Langerin favors CD8<sup>+</sup> T cell cellular (T<sub>H</sub>1) responses, whereas targeting of DCIR2 favors CD4<sup>+</sup> T cell and B cell humoral (T<sub>H</sub>2) responses (Cruz et al., 2012). Vaccine potency may be maximized by targeting multiple DC subsets, thereby inducing both cellular and humoral immune responses (Ueno et al., 2011).

In addition to delivering information regarding specificity and activity to DCs, some investigators have considered the design of artificial antigen-presenting cells (aAPCs) that can cross-prime antigen-specific CD8<sup>+</sup> T cells directly. Synthetic aAPCs are particles to which proteins required for T cell activation—such as MHC-epitope or agonist anti-CD3 (Signal 1 to the T cell receptor) and agonist anti-CD28 (co-stimulatory Signal 2)—have been conjugated. Manipulating particle shape and geometry revealed that aAPC activity correlates with aspect ratio (Sunshine et al., 2014). Mechanistically, CD8<sup>+</sup> T cells migrate preferentially to the long axis of ellipsoidal aAPCs, and this extended length of contact increases T cell proliferation and, consequently, tumor prevention.

These data not only have relevance to design parameters for future aAPCs but also provide insights into the fundamental biology of DC-T cell interactions. Indeed, the result was

CD8<sup>+</sup> T cells, which are critical mediators of antitumor immunity. Sustained antigen release from a particle within DCs can further enhance cytolytic T lymphocyte (CTL) priming in vitro by extending antigen presentation (Audran et al., 2003). Such particles serve as antigen reservoirs, thereby mimicking both prime and boost injections following a single administration.

As yet, the induction of robust CD8<sup>+</sup> T cell responses in large animal models has not been achieved using traditional protein vaccine-based approaches, which have primarily elicited humoral B cell responses. Excitingly, preliminary studies suggest that nanoparticle-based vaccines may confer cross-priming efficacy in non-human primates and humans similar to that observed in mice. Should such findings be validated in larger cohorts, then nanovaccines would serve as an important breakthrough for the development of vector-free vaccines (Irvine et al., 2013).

Owing to their pathogen-like size, nanoparticles are readily taken up by antigen-presenting cells, such as DCs, which are natural phagocytes. As a consequence, even untargeted nanoparticles improve the uptake—which often correlates with antitumor efficacy—of cancer vaccines relative to their soluble forms. Altering particle size, hydrophobicity, and surface charge, as well as conjugating targeting ligands, can further enhance uptake efficiency (Cruz et al., 2012). Targeting nanoparticles to DCs has proven much more feasible than targeting nanoparticles to cancer cells. This difference is rooted in physics as much as in

unexpected, as high aspect ratio has previously been associated with increased particle internalization by non-phagocytic cells. In addition to nanoparticles, scaffolds can be used to vaccinate against cancer. Scaffolds similarly offer practical and functional advantages over conventional DC-based vaccines, which require isolation, ex vivo manipulation, and reintroduction of a patient's DCs.

### Scaffolds: Customized Microenvironments

Polymeric scaffolds and hydrogels can be implanted or injected to generate a modular, tailored local microenvironment that can co-localize inflammatory cytokines, tumor antigen, and immune danger signals in situ. For example, incorporation of the chemokine GM-CSF, autologous tumor lysate, and the TLR9 agonist CpG-oligonucleotide into a subcutaneously implanted porous PLGA scaffold promotes recruitment and activation of DCs, resulting in regression of established local and distant tumors (Ali et al., 2009). The observed potency is attributed to the persistent presence of antigen and adjuvant signaling in the depot, which is not attainable upon delivery of soluble vaccine components that diffuse away fairly rapidly. This exciting scaffold-based approach is currently being investigated in a phase I clinical trial (NCT01753089). To avoid the need for surgical implantation, an injectable spontaneously assembling scaffold has been devised (Kim et al., 2015). Specifically, mesoporous silica rods with high aspect ratios form macroporous structures that provide a favorable microenvironment for DCs, which subsequently traffic to lymph nodes and provoke adaptive immune responses.

Scaffolds can similarly be used to improve the function of adoptively transferred T cells by providing a supportive immunologic microenvironment. Adoptive cell transfer (ACT), particularly upon introduction of chimeric antigen receptors into T cells (Maude et al., 2014), can lead to sustained remissions in hematologic malignancies. Solid tumors, however, establish a concentrated immunosuppressive microenvironment that hampers the efficacy of ACT. Transplantation of lymphocytes in biodegradable polymeric scaffolds can sustain expansion and release of tumor-reactive T cells at tumor resection sites and enhance their antitumor potency (Stephan et al., 2015). Scaffold-derived T cells reduce residual disease and relapse much more effectively than free T cells administered systemically or locally. Such depots provide proof of concept for localized delivery of cells, in addition to small molecules and biologics. Localized immunotherapy is particularly well suited for treatment of inoperable or incompletely removed tumors to prevent local recurrence (Stephan et al., 2015), and its effects can have widespread implications.

### Localized Nanoimmunotherapy: Focal Impact

Achieving a robust local antitumor effect—as previously observed for the radiation-induced “abscopal effect”—can generate a systemic antitumor immune response that can eradicate disseminated disease, including metastases situated in sites generally thought to be tumor cell havens in the context of traditional systemic therapy (Marabelle et al., 2013). Liposomes can be used to anchor immunomodulatory compounds, such as immunostimulatory nucleic acids and biologics (Kwong et al., 2013), prior to intratumoral injection. Such particles restrict

the biodistribution of these compounds and prolong their retention at the tumor site. In so doing, localized nanoimmunotherapy reduces systemic toxicity and thus improves the therapeutic window of extremely potent immunostimulatory molecules while still promoting systemic antitumor immunity (Kwong et al., 2011).

### Delivering Tx to T Cells in Circulation

Immunoengineering also enables drug delivery directly to T cells. The conjugation of nanoparticles loaded with supportive compounds to the surface of adoptively transferred T cells leads to persistent autocrine-like signaling among these “pharmacytes” (Stephan et al., 2010). This approach again demonstrates the impact of nanotechnology relative to administration of free drug and represents a paradigm that can be applied more broadly than the ACT-supportive scaffold described above, as it does not necessitate surgical implantation. Ideally, one would be able to deliver such adjuvant drug-containing nanoparticles to T cells upon systemic administration, enabling a generalized approach that does not require ex vivo cell manipulation for each patient. Excitingly, liposomes to which targeting ligands—antibody fragments or cytokines—have been conjugated can target drug delivery to adoptively transferred T cells in vivo (Zheng et al., 2013). Future work will likely enable targeted delivery to endogenous T cells and, ultimately, other cell types as well.

### Concentrating Catalysis

In addition to delivering small molecules, oligonucleotides, antigens, and cytokines, nanoparticles can be used to concentrate enzymes in vivo. For example, particles can be used to degrade neutrophil extracellular traps (NETs). NETs are extracellular DNA structures that, when formed intravascularly, can sequester circulating tumor cells and thereby promote metastasis. Digesting NETs with free DNase is relatively inefficient at inhibiting metastasis (Cools-Lartigue et al., 2013), while DNase-coated nanoparticles vastly improve therapeutic efficacy (J. Park, R.W. Wysocki, Z. Amoozgar, M.S.G., and M. Egeblad, unpublished data).

### Looking Ahead

Moving forward, the field of immunoengineering will benefit from a broader adoption of novel tools that permit multiplex analysis of cell type (multiplexed ion beam imaging), cell activation state (mass cytometry), and soluble mediators of stimulation/inhibition (Luminex) in the tumor microenvironment and circulation following perturbation. By allowing for interrogation of several-fold more parameters simultaneously than conventional methodologies, such as flow cytometry and ELISA, such tools will yield insights into the coordination of the highly complex immune system. A comprehensive understanding of the downstream impacts of our interventions, including expression of co-stimulatory/inhibitory ligands and production of immunoregulatory cytokines, will enable rational product revision for improved therapeutic outcomes.

Beyond enhancing our appreciation of the cellular and biochemical constituents of the tumor microenvironment, we will benefit from an increased consideration of the physical microenvironment in the tumor, as well as its draining lymph nodes (Swartz and Lund, 2012). Indeed, extracellular matrix serves as a physical mediator of immunosuppression by preventing

penetration of immune cells into the tumor core (Salmon and Donnadieu, 2012). Immunoengineering can be used to alter the physical microenvironment of tumors, for example, by modifying peritumoral extracellular matrix (Kanapathipillai et al., 2012).

To this end, in addition to enabling more thorough descriptions of immune cell function (reading), advanced technologies can be used to create physical lymphoid-like structures to study and manipulate immune cell function (writing). Improving the reproducibility of formulation and fabrication methods is critical, as manufacturing represents perhaps the greatest obstacle confronting the clinical translation of nanodevices. Controlled production can minimize polydispersity. For example, 3D printing is revolutionizing the field of regenerative medicine, and it has the potential to influence cancer immunoengineering similarly. Initially, this technology will likely be applied to produce defined, improved scaffolds for vaccine applications or supportive ACT. In the years ahead, it could be used to create implantable artificial tertiary lymphoid structures, which possess defined zones for specialized immune cells and are important to long-term cancer patient survival (Fridman et al., 2012).

Transplantable lymphoid-like organoids can already be engineered to manifest discrete compartments for particular immune cells, which generate functional humoral and cellular responses to vaccination (Suematsu and Watanabe, 2004). 3D printing will allow for deposition of specific cytokines, immune cells, and matrix with unprecedented accuracy. This advance will have relevance not only to translational biology but also to basic immunology, as engineered scaffolds can enhance our understanding of the biochemical and physical microenvironments that alter the balance between tolerance and rejection (Swartz et al., 2012).

By concentrating the delivery of their payloads, nanoparticles permit the use of considerably lower doses of immunostimulatory molecules to achieve a given response and thereby enhance the safety profiles of these drugs (Irvine et al., 2013). Still, the materials from which nanodevices are created can inherently provoke a host response, so meaningful safety parameters must be defined, such as serum levels of type I interferons and IL-6. Unlike for prophylactic vaccines, such responses are likely acceptable to cancer patients and may even be beneficial in stimulating antitumor immunity, but they must be well understood nonetheless. Encouragingly, tocilizumab (anti-IL6R) has been used to manage cytokine-release syndrome successfully in the acute setting (Maude et al., 2014). Placing an emphasis on the development of safe biomaterials will facilitate earlier translation of immunoengineered products into patients. Data gleaned from patients will be more informative than anything that can be derived from preclinical models.

Emerging evidence confirms that cancer immunotherapies, which can generate adaptive and durable responses, yield much more robust antitumor effects when they are formulated in nanoparticles or scaffolds than when they are administered as free drugs. Cancer immunoengineering is thus a promising area worthy of further consideration and investigation. It is hoped that this piece will stimulate basic biologists to engage bioengineers and to articulate the questions that they would like to see

addressed with innovative technologies. In addition to its therapeutic potential, immunoengineering provides a valuable tool for dissecting fundamental biology.

## ACKNOWLEDGMENTS

The author thanks the Cancer Research Institute, Melanoma Research Alliance, and Ovarian Cancer Research Fund for their support and apologizes to those investigators whose work could not be referenced because of space limitations.

## REFERENCES

- Ali, O.A., Emerich, D., Dranoff, G., and Mooney, D.J. (2009). *Sci. Transl. Med.* 1, 8ra19.
- Audran, R., Peter, K., Dannull, J., Men, Y., Scandella, E., Groettrup, M., Gander, B., and Corradin, G. (2003). *Vaccine* 21, 1250–1255.
- Chow, E.K., and Ho, D. (2013). *Sci. Transl. Med.* 5, 216rv214.
- Cools-Lartigue, J., Spicer, J., McDonald, B., Gowing, S., Chow, S., Giannias, B., Bourdeau, F., Kubes, P., and Ferri, L. (2013). *J. Clin. Invest.* 123, 3446–3458.
- Cruz, L.J., Tacke, P.J., Rueda, F., Domingo, J.C., Albericio, F., and Figdor, C.G. (2012). *Methods Enzymol.* 509, 143–163.
- Fridman, W.H., Pagès, F., Sautès-Fridman, C., and Galon, J. (2012). *Nat. Rev. Cancer* 12, 298–306.
- Irvine, D.J., Swartz, M.A., and Szeto, G.L. (2013). *Nat. Mater.* 12, 978–990.
- Kanapathipillai, M., Mammoto, A., Mammoto, T., Kang, J.H., Jiang, E., Ghosh, K., Korin, N., Gibbs, A., Mannix, R., and Ingber, D.E. (2012). *Nano Lett.* 12, 3213–3217.
- Kim, J., Li, W.A., Choi, Y., Lewin, S.A., Verbeke, C.S., Dranoff, G., and Mooney, D.J. (2015). *Nat. Biotechnol.* 33, 64–72.
- Kwong, B., Liu, H., and Irvine, D.J. (2011). *Biomaterials* 32, 5134–5147.
- Kwong, B., Gai, S.A., Elkhader, J., Wittrup, K.D., and Irvine, D.J. (2013). *Cancer Res.* 73, 1547–1558.
- Marabelle, A., Kohrt, H., Sagiv-Barfi, I., Ajami, B., Axtell, R.C., Zhou, G., Rajapaksa, R., Green, M.R., Torchia, J., Brody, J., et al. (2013). *J. Clin. Invest.* 123, 2447–2463.
- Maude, S.L., Frey, N., Shaw, P.A., Aplenc, R., Barrett, D.M., Bunin, N.J., Chew, A., Gonzalez, V.E., Zheng, Z., Lacey, S.F., et al. (2014). *N. Engl. J. Med.* 371, 1507–1517.
- Salmon, H., and Donnadieu, E. (2012). *Oncol Immunology* 1, 992–994.
- Stephan, M.T., Moon, J.J., Um, S.H., Bershteyn, A., and Irvine, D.J. (2010). *Nat. Med.* 16, 1035–1041.
- Stephan, S.B., Taber, A.M., Jileeva, I., Pegues, E.P., Sentman, C.L., and Stephan, M.T. (2015). *Nat. Biotechnol.* 33, 97–101.
- Suematsu, S., and Watanabe, T. (2004). *Nat. Biotechnol.* 22, 1539–1545.
- Sunshine, J.C., Perica, K., Schneck, J.P., and Green, J.J. (2014). *Biomaterials* 35, 269–277.
- Swartz, M.A., and Lund, A.W. (2012). *Nat. Rev. Cancer* 12, 210–219.
- Swartz, M.A., Hirose, S., and Hubbell, J.A. (2012). *Sci. Transl. Med.* 4, 148rv149.
- Topalian, S.L., Weiner, G.J., and Pardoll, D.M. (2011). *J. Clin. Oncol.* 29, 4828–4836.
- Ueno, H., Klechevsky, E., Schmitt, N., Ni, L., Flamar, A.L., Zurawski, S., Zurawski, G., Palucka, K., Banchereau, J., and Oh, S. (2011). *Semin. Immunol.* 23, 21–27.
- Zheng, Y., Stephan, M.T., Gai, S.A., Abraham, W., Shearer, A., and Irvine, D.J. (2013). *J. Control. Release* 172, 426–435.



# Immune Checkpoint Targeting in Cancer Therapy: Toward Combination Strategies with Curative Potential

Padmanee Sharma<sup>1,2,\*</sup> and James P. Allison<sup>1,\*</sup>

<sup>1</sup>Department of Immunology

<sup>2</sup>Department of Genitourinary Medical Oncology

MD Anderson Cancer Center, Houston, TX 77030, USA

\*Correspondence: [padsharma@mdanderson.org](mailto:padsharma@mdanderson.org) (P.S.), [jallison@mdanderson.org](mailto:jallison@mdanderson.org) (J.P.A.)

<http://dx.doi.org/10.1016/j.cell.2015.03.030>

Research in two fronts has enabled the development of therapies that provide significant benefit to cancer patients. One area stems from a detailed knowledge of mutations that activate or inactivate signaling pathways that drive cancer development. This work triggered the development of targeted therapies that lead to clinical responses in the majority of patients bearing the targeted mutation, although responses are often of limited duration. In the second front are the advances in molecular immunology that unveiled the complexity of the mechanisms regulating cellular immune responses. These developments led to the successful targeting of immune checkpoints to unleash anti-tumor T cell responses, resulting in durable long-lasting responses but only in a fraction of patients. In this Review, we discuss the evolution of research in these two areas and propose that intercrossing them and increasing funding to guide research of combination of agents represent a path forward for the development of curative therapies for the majority of cancer patients.

## Introduction

The scientific community united against a common enemy in 1971 when President Nixon signed a bill initiating the “War on Cancer,” which provided funding for scientific research focused on improving our understanding and treatment of cancer. Without doubt, the intervening years were followed by great advances in the elucidation of the molecular mechanisms that regulate growth and death of normal cells, including a deep understanding of how these pathways progressively go awry during the development of cancer. This understanding led to the era of genomically targeted therapies and “precision medicine” in the treatment of cancer. Genomically targeted therapies can result in remarkable clinical responses. The ability of cancer cells to adapt to these agents by virtue of their genomic instability and other resistance mechanisms eventually leads to disease progression in the majority of patients nonetheless. Unraveling the mechanisms by which cancer cells become resistant to drugs and developing new agents to target the relevant pathways have become logical next steps in this approach for cancer treatment. However, given the genetic and epigenetic instability of cancer cells, it is likely that each new drug or combination of drugs targeting the tumor cells will meet with more complex mechanisms of acquired resistance. Recent findings suggest that T cells, bearing antigen receptors that are generated by random rearrangement of gene segments, followed by selective processes that result in a vast repertoire of T cell clones, provide sufficient diversity and adaptability to match the complexity of tumors. Discoveries regarding regulation of T cell responses have provided key principles regarding immune

checkpoints that are being translated into clinical success, with durable responses and long-term survival greater than 10 years in a subset of patients with metastatic melanoma, as well as yielding promising results in several other tumor types. Now, with the perspective of combining genomically targeted agents and immune checkpoint therapies, we are finally poised to deliver curative therapies to cancer patients. To support this goal and accelerate these efforts, changes in directions of research support and funding may be required.

## Precision Medicine: Targeting the Drivers

In the past three decades, enormous strides have been made in elucidating the molecular mechanisms involved in the development of cancer (Hanahan and Weinberg, 2011). It is now clear that the oncogenic process involves somatic mutations that result in activation of genes that are normally involved in regulation of cell division and programmed cell death, as well as inactivation of genes involved in protection against DNA damage or driving apoptosis (Bishop, 1991; Solomon et al., 1991; Weinberg, 1991; Knudson, 2001). These genetic links led to the decision early in the war on cancer to undertake sequencing of cancer genomes to provide a comprehensive view of somatic mutational landscapes in cancer and identify possible therapeutic targets. Infrastructure and funding were provided to coordinate the sequencing efforts. It has become apparent that the level of somatic mutations differs widely between and within different tumor types ranging from very low rates in childhood leukemias to very high rates in tumors associated with carcinogens (Alexandrov et al., 2013).

Mutations can be divided into two broad classes: those whose products “drive” tumorigenesis in a dominant fashion and “passengers” with no obvious role in the tumor causation. The Cancer Genome Atlas (TCGA) projects have enabled identification of many of these mutations (Chen et al., 2014; Cancer Genome Atlas Research Network, 2014). This has allowed for the rational design of drugs that target and selectively interfere with oncogenic signaling pathways. This approach has revolutionized cancer medicine by moving away from the “one size fits all” approach—for instance, traditional chemotherapy, which attacks all dividing cells, including both cancer-differentiating or regenerating normal cells—to a more personalized strategy of treating patients with a specific drug only if their cancer bears particular molecular mutations that are target of that drug.

As an example of genomically targeted therapies, an inhibitor against BRAF was developed when it was discovered that ~40%–60% of cutaneous melanomas carry mutations in BRAF, which induces constitutive activation of the MAPK pathway (Curtin et al., 2005; Davies et al., 2002). In a randomized phase III trial comparing a BRAF inhibitor (vemurafenib) versus dacarbazine, the vemurafenib treatment group had a response rate of ~48% versus 5% in the dacarbazine arm (Chapman et al., 2011). However, the median duration of response was short, only 6.7 months (Sosman et al., 2012). Another oncogenic pathway that has been targeted is the tyrosine kinase chromosomal rearrangement, which results in the fusion oncogene EML4-ALK that is found in ~5% of NSCLC patients (Soda et al., 2007). The EML4 fusion partner mediates ligand-independent oligomerization and/or dimerization of anaplastic lymphoma kinase (ALK), resulting in constitutive kinase activity. Standard chemotherapies in this subgroup of patients have been associated with response rates of up to 10% (Hanna et al., 2004). Crizotinib, a tyrosine kinase inhibitor targeting ALK (Kwak et al., 2010), was shown to elicit a response rate of ~65% with a median duration of response of less than 8 months in a phase III trial (Shaw et al., 2013). Although there was a significant increase in progression-free survival for patients treated with crizotinib, regrettably, there was no overall survival benefit in the interim analysis. Therefore, although the concept of targeting “driver mutations” has great merit and has demonstrated clinical responses, the reality remains that the majority of patients treated with these agents will derive short-term clinical responses with eventual development of resistance mechanisms that lead to disease progression and death.

Mechanisms operative in acquired resistance fall into three main categories: alterations in the targeted gene (as a result of mutation, amplification, or alternative splicing); other changes that do not affect the original target but re-activate the signaling pathway involved (i.e., NRAS and MEK mutations in BRAF mutant melanoma); and changes that activate alternate pathways (such as activation of growth factor receptors). Considerable effort has gone into finding ways to enhance efficacy of genomically targeted therapies. One effort involves multiple agents that target different molecules in the same pathway, such as the combination of a BRAF inhibitor and a MEK-inhibitor (Larkin et al., 2014; Robert et al., 2015a). This approach helps to reduce compensatory feedback loops, as well as to block the development of resistance due to mutations downstream that

pathway. A different strategy consists of blocking parallel pathways to prevent emerging resistance (Martz et al., 2014). Still, the chief challenge of these combinatorial approaches is the multiplicity of resistance mechanisms and the fact that different mechanisms may be in operation in different cells due to intratumor heterogeneity. Given these observations, it is difficult to envision realistic approaches to effectively overcome the myriad of resistance mechanisms that may arise in the course of cancer treatment. The continued evolvability of the tumor cells and their mechanisms of escape from targeted therapies raise the question as to whether combinations of genomically targeted agents will ever be curative.

### Advantages of Mobilizing T Cells for Cancer Therapy

As the knowledge of the intricate biology of cancer has progressed, so has the understanding of the fundamental cellular and molecular mechanisms that orchestrate the interplay of the innate and adaptive arms of the immune system. In a simplistic way, the innate system is composed primarily of cytokines, the complement system, and phagocytes such as macrophages, neutrophils, dendritic cells, and natural killer (NK) cells. Cells of the innate immune system have hard-wired receptors to detect products of infectious microorganisms and dying cells. Macrophages and neutrophils provide an early defense against microorganisms, whereas dendritic cells provide a key interface to the adaptive immune system, composed of B and T cells with their somatically generated, clonally expressed repertoire of antigen receptors.

The understanding of the basic principles governing the controlling immunity provided the rational for the development of powerful strategies to actively engage the immune system for cancer therapy. Strategies to unleash T cells against tumors are particularly compelling, as the activity of these cells presents important features that are advantageous over other cancer therapies. The first is their specificity. T cells express antigen receptors that recognize cell-surface complexes of MHC molecules and peptides sampled from virtually all the proteins in the cell and are not limited to peptide antigens derived from cell-surface molecules. The second feature is memory. Primary T cell responses are generally followed by the production of long-lived memory T cells with accelerated kinetics of secondary response if the antigen recurs. Finally, the T cell response is adaptable and can accommodate not only tumor heterogeneity but also responses to novel antigens expressed by recurring tumors. It has been calculated that the somatic recombination process that generates the antigen receptors of T cells can generate as many as  $10^{15}$  different receptors (Davis and Bjorkman, 1988). Of this theoretical number, each individual human has perhaps  $10^9$  different receptors. The immense size of the repertoire suggests that the immune system is indeed well equipped to deal with mutability and adaptability of cancer.

### Harnessing T Cell Responses to Tumor Antigens

With the advent of genomic and cDNA expression cloning methods and sequencing of peptides eluted from tumor cell MHC molecules, an avalanche of tumor antigens defined by tumor-specific T cells has been identified in both mice and in humans. Most of these are shared between cancer cells of different

individuals and fall into four groups: products of oncogenic viruses (Epstein-Barr virus in certain leukemias and human papilloma virus in cervical and some head and neck cancers); antigens related to tissue-specific differentiation molecules (tyrosinase and related proteins in melanoma and prostate-specific antigen and prostatic acid phosphatase in prostate cancer); molecules normally expressed only during fetal development (carcino-embryonic antigen in colon cancer,  $\alpha$ -fetoprotein in liver cancer); and cancer-testes (CT) antigens, which are normally expressed during gametogenesis but are found in many cancer cells as a result of changes in epigenetic regulation (MAGE and NY-ESO-1).

Additionally, somatic mutations also can result in the generation of tumor-specific peptides with the potential to bind major histocompatibility complex (MHC) molecules and therefore be recognized by the immune system as neoantigens (Sjöblom et al., 2006; Segal et al., 2008). The analysis of the epitope landscape of breast and colon carcinoma cells revealed that the products of seven to ten mutant genes in colorectal and breast cancer, respectively, have the potential for binding to HLA-A\*0201 alone. Because each heterozygote individual carries as many as 6 different HLA class I genes, this means an average of 42–60 potential neoantigens that can be presented to T cells. In support of these estimates, recent studies have demonstrated that neoantigens generated by somatic mutation are recognized by T cells in both mouse and human cancers (Linemann et al., 2015; Gros et al., 2014; Tran et al., 2014; Gubin et al., 2014).

At first, as a result of earlier studies identifying shared antigens, the field of cancer immunotherapy became focused on developing therapeutic vaccines to expand T cells against these shared antigens expressed on tumors. Many studies focused on stimulating T cell responses with peptides, proteins, whole-tumor cells including those modified to express cytokines, DNA, recombinant viral-based vaccines, or antigen-pulsed dendritic cells given alone or in combination with various adjuvants or cytokines. Although these trials were conducted with the best available science at the time and provided promising anecdotal evidence that induction of immune responses could elicit clinical benefit, they remained largely negative and generally failed to show objective clinical responses (see Rosenberg et al., 2004 for review). Enthusiasm waned somewhat as the number of failed clinical trials mounted.

Many reasons might have contributed to the failure of these vaccination strategies, including choice of antigen, failure to provide adequate costimulation, or functional inactivation of tumor-reactive T cells (Melero et al., 2014). A number of T-cell-extrinsic suppressive mechanisms such as TGF $\beta$ , FoxP3<sup>+</sup> regulatory T cells (Treg), and tryptophan metabolites (IDO) that can hamper anti-tumor responses have also been identified, and there have been efforts to minimize the suppressive effects of these in pre-clinical and clinical studies.

### Unraveling the Complexity of T Cell Activation

Another contributing factor to the failure of earlier cancer vaccine trials was perhaps the lack of understanding and appreciation of the full complexity of cell-intrinsic pathways that regulate T cell activation. By the late 1980s, it was known that simple engage-

ment of peptide/MHC complexes by the antigen receptor is insufficient for activation of T cells and may render them anergic (Jenkins and Schwartz, 1987; Mueller et al., 1989). In order to become fully activated, T cells must encounter antigen in the context of antigen-presenting cells (APCs) such as dendritic cells, which provide costimulatory signals mediated by B7 molecules (B7-1 and B7-2) that will engage their ligand, CD28, in the T cell (Greenwald et al., 2005). Thus, T cells specific for a tumor antigen will not be activated by an initial encounter with tumor cells or may even be rendered anergic because, with the exception of a few lymphomas, tumors do not express costimulatory B7 molecules (Townsend and Allison, 1993). Thus, tumors are essentially invisible to T cells until the T cells are activated as a result of cross-priming by dendritic cells that present tumor antigens acquired from dying tumor cells. Simultaneous recognition of antigen/MHC complexes and costimulatory ligands by T cells initiates a complex set of genetic programs that result in cytokine production, cell-cycle progression, and production of anti-apoptotic factors that result in proliferation and functional differentiation of T cells. Consistent with the importance of both antigen receptor and costimulatory signals in initiating anti-tumor responses, many therapeutic vaccines now incorporate both antigen and dendritic cells or agents that enhance costimulatory signaling.

By the mid-90s, it became clear that T cell priming elicits not only programs leading to induction of T cell responses but also a parallel program that will eventually stop the response. The critical inhibitory program is mediated by CTLA-4, a homolog of CD28 that also binds B7-1 and B7-2, although with much greater avidity than that CD28. Expression of the *ctla-4* gene is initiated upon T cell activation, and it traffics to and accumulates in the immunological synapse, eventually attenuating or preventing CD28 costimulation by competition for B7 binding and negative signaling (Walunas et al., 1994; Krummel and Allison, 1995). The fact that *ctla-4* knockout mice suffer from a rapid and lethal lymphadenopathy (Waterhouse et al., 1995; Tivol et al., 1995; Chambers et al., 1997) speaks for a negative role for CTLA-4 in limiting T cell responses to prevent damage to normal tissues.

Thus, activation of T cells as a result of antigen receptor signaling and CD28 costimulation is followed not only by induction of genetic programs leading to proliferation and functional differentiation but also by induction of an inhibitory program mediated by CTLA-4, which will ultimately stop proliferation. Extrapolating this paradigm to anti-tumor T cell responses, if eradication of the tumor has not been completed by the time that the inhibitory signal of CTLA-4 is triggered, the T cells will be turned off and will be unable to complete the task. Importantly, this also suggests that, after this program is initiated, vaccines used to stimulate antigen receptor signaling may actually serve to strengthen the “off” signal as a result of additional induction of *ctla-4* expression by antigen receptor signaling. In any event, this suggests the importance of shifting strategies for cancer immunotherapy from activating T cells to unleashing them.

### Inactivating the Brakes to Increase Anti-tumor Immunity

Consistent with the observations that CD28 and CTLA-4 had opposing effects on T cell responses in vitro, in the late 90s, it

was found that, although blocking antibodies to CD28 impaired anti-tumor responses in mice, blocking antibodies to CTLA-4 enhanced anti-tumor responses in mouse tumor models (Leach et al., 1996). In fact, the treatment of mice with anti-CTLA-4 antibodies as monotherapy results in complete tumor rejection and long-lived immunity. Later on, mechanistic studies revealed that anti-tumor activity was associated with increased ratio of both CD4 and CD8 effector cells to FoxP3<sup>+</sup> regulatory T cells (Quezada et al., 2006). The success of CTLA-4 blockade in these initial studies raised two compelling points. First, because the target molecule was on the T cell and not the tumor cell, it was feasible to imagine that the same strategy would work on many different histologic tumors, as well as on tumors caused by different genetic lesions. Second, taking into consideration that CTLA-4 inhibited CD28-mediated costimulation by a cell-intrinsic mechanism (Peggs et al., 2009), its blockade could allow for enhanced T cell costimulation, which in turn would increase the efficacy of tumor vaccines, as well as agents that kill tumor cells under conditions that promote inflammatory responses. These possibilities were further supported by the results of a series of studies in different mouse models, including the demonstration that blockade of CTLA-4 was not limited to any particular tumor type but was rather broadly effective. CTLA-4 also was able to synergize with a vaccine consisting of tumor cells engineered to express the cytokine GM-CSF to eradicate tumors (Hurwitz et al., 1998; van Elsland et al., 1999). Finally, CTLA-4 could be combined with local delivery of irradiation, cryoablation, or an oncolytic virus to induce systemic tumor immunity and eradication of distant metastases (Zamarin et al., 2014; Waitz et al., 2012; Tang et al., 2014). These preclinical studies supported the development of clinical anti-CTLA-4 therapy.

### Immune Checkpoint Therapy: The Clinical Success

CTLA-4 blockade was translated to the clinic with a fully human antibody to human CTLA-4 (ipilimumab, Medarex, Bristol-Myers Squibb). Tumor regression was observed in phase I/II trials in patients with a variety of tumor types, including melanoma, renal cell carcinoma, prostate cancer, urothelial carcinoma, and ovarian cancer (Yang et al., 2007; Hodi et al., 2008; Carthon et al., 2010; van den Eertwegh et al., 2012). Two phase III clinical trials with ipilimumab were recently completed in prostate cancer, the first in patients with castrate-resistant prostate cancer who had not received prior chemotherapy treatment and the second in a more advanced disease setting, in which patients with castrate-resistant prostate cancer presented disease that had progressed on chemotherapy treatment. The former trial is yet to be reported. The latter trial reports the lack of statistical significance (*p* value of 0.053) to indicate a survival benefit for patients who received ipilimumab treatment. However, subset analyses indicate that patients who have favorable clinical characteristics such as lack of liver metastases do benefit from ipilimumab therapy (Kwon et al., 2014). Two phase III clinical trials with anti-CTLA-4 (ipilimumab) were also conducted in patients with advanced melanoma and demonstrated improved overall survival for patients treated with ipilimumab (Hodi et al., 2010; Robert et al., 2011). Importantly, these trials indicate long-term durable responses with greater than 20% of treated patients

living for more than 4 years, including a recent analysis indicating survival of 10 years or more for a subset of patients (Schadendorf et al., 2015). The FDA approved ipilimumab as treatment for patients with melanoma in 2011.

The clinical success of anti-CTLA-4 opened a new field termed “immune checkpoint therapy” as additional T cell intrinsic pathways were identified and targeted for clinical development (Sharma et al., 2011; Pardoll, 2012). Another T-cell-intrinsic inhibitory pathway identified after CTLA-4 was that mediated by PD-1 (programmed death 1) and its ligand PD-L1. PD-1 was initially cloned in 1992 in a study of molecules involved in negative selection of T cells by programmed cell death in the thymus (Ishida et al., 1992). Its function as an immune checkpoint was not established until 2000 upon identification of its ligands (Freeman et al., 2000). PD-L1 was then shown to protect tumor cells by inducing T cell apoptosis (Dong et al., 2002). Later, preclinical studies in animal models evaluated anti-PD-1 and anti-PD-L1 antibodies as immune checkpoint therapies to treat tumors (Keir et al., 2008).

Much like CTLA-4, PD-1 is expressed only in activated T cells. However, unlike CTLA-4, PD-1 inhibits T cell responses by interfering with T cell receptor signaling as opposed to outcompeting CD28 for binding to B7. PD-1 also has two ligands, PD-L1 and PD-L2. PD-L2 is predominantly expressed on APCs, whereas PD-L1 can be expressed on many cell types, including cells comprising the immune system, epithelial cells, and endothelial cells. Antibodies targeting PD-L1 have shown clinical responses in multiple tumor types, including melanoma, renal cell carcinoma, non-small-cell lung cancer (Brahmer et al., 2012), and bladder cancer (Powles et al., 2014). Similarly, phase I clinical trials with a monoclonal antibody against PD-1 demonstrated clinical responses in multiple tumor types, including melanoma, renal cell carcinoma, non-small-cell carcinoma (Topalian et al., 2012), Hodgkin's lymphoma (Ansell et al., 2015), and head and neck cancers (Seiwert et al., 2014, J. Clin. Oncol., abstract). Recently, a large phase I clinical trial with an anti-PD-1 antibody known as MK-3475 showed response rates of ~37%–38% in patients with advanced melanoma, including patients who had progressive disease after prior ipilimumab treatment (Hamid et al., 2013), triggering the approval of MK-3475 (pembrolizumab, Merck) by the FDA in September 2014. A phase III clinical trial that treated patients with metastatic melanoma with a different anti-PD-1 antibody (nivolumab, Bristol-Myers Squibb, BMS) also demonstrated improved responses and overall survival benefit as compared to chemotherapy treatment (Robert et al., 2015b). Nivolumab was FDA approved for patients with metastatic melanoma in December 2014. In addition, nivolumab was FDA approved in March 2015 for patients with previously treated advanced or metastatic non-small-cell lung cancer based on a phase III clinical trial, which reported an improvement in overall survival for patients treated with nivolumab as compared to patients treated with docetaxel chemotherapy.

Because CTLA-4 and PD-1 regulate different inhibitory pathways on T cells, combination therapy with antibodies targeting both molecules was tested and found to improve anti-tumor responses in a pre-clinical murine model (Curran et al., 2010). A recently reported phase I clinical trial with anti-CTLA-4 in combination with anti-PD-1 also demonstrated tumor regression



in ~50% of treated patients with advanced melanoma, in most cases with tumor regression of 80% or higher (Wolchok et al., 2013). There are ongoing clinical trials with anti-CTLA-4 (ipilimumab, BMS or tremelimumab, MedImmune/Astrazeneca) plus anti-PD-1 or anti-PD-L1 in other tumor types, with preliminary data indicating promising results (Hammers et al., 2014, J. Clin. Oncol., abstract; Callahan et al., 2014, J. Clin. Oncol., abstract) that highlight this combination as an effective immunotherapy strategy for cancer patients.

As with other cancer therapies, immune checkpoint therapies may lead to side effects and toxicities (see Postow et al., 2015; Gao et al., 2015 for recent reviews). Briefly, these side effects consist of immune-related adverse events that are defined by inflammatory conditions, including dermatitis, colitis, hepatitis, pancreatitis, pneumonitis, and hypophysitis. These side effects can be managed and usually involve administration of immunosuppressive agents such as corticosteroids, which do not appear to interfere with clinical benefit that is derived from the immune checkpoint agents. The profile of side effects that occur with both anti-CTLA-4 and anti-PD-1/PD-L1 antibodies is similar; however, the side effects appear to occur more frequently in the setting of anti-CTLA-4 therapy as compared to anti-PD-1 and anti-PD-L1 therapies. The continued success of immune checkpoint therapies in the clinic will require education of the oncology community regarding recognition and treatment of the side effects elicited by these agents.

### Novel Immunologic Targets for Cancer Immunotherapy

Although blockade of the CTLA-4 and PD-1/PD-L1 pathways is furthest along in clinical development, it only represents the tip of the iceberg in the realm of potential targets that can serve to improve anti-tumor responses. Ongoing studies on regulation of immune responses have led to the identification of multiple other immunologic pathways that may be targeted for the development of therapies, either as monotherapy or in combination strategies, for the successful treatment of cancer patients. These include immune checkpoints or inhibitory pathways, as well as co-stimulatory molecules, which act to enhance immune responses. A partial list of new immune checkpoints that are being evaluated in pre-clinical tumor models and/or in the clinic with cancer patients includes LAG-3 (Triebel et al., 1990), TIM-3 (Sakuishi et al., 2010), and VISTA (Wang et al., 2011), whereas co-stimulatory molecules include ICOS (Fan et al., 2014), OX40 (Curti et al., 2013), and 4-1BB (Melero et al., 1997).

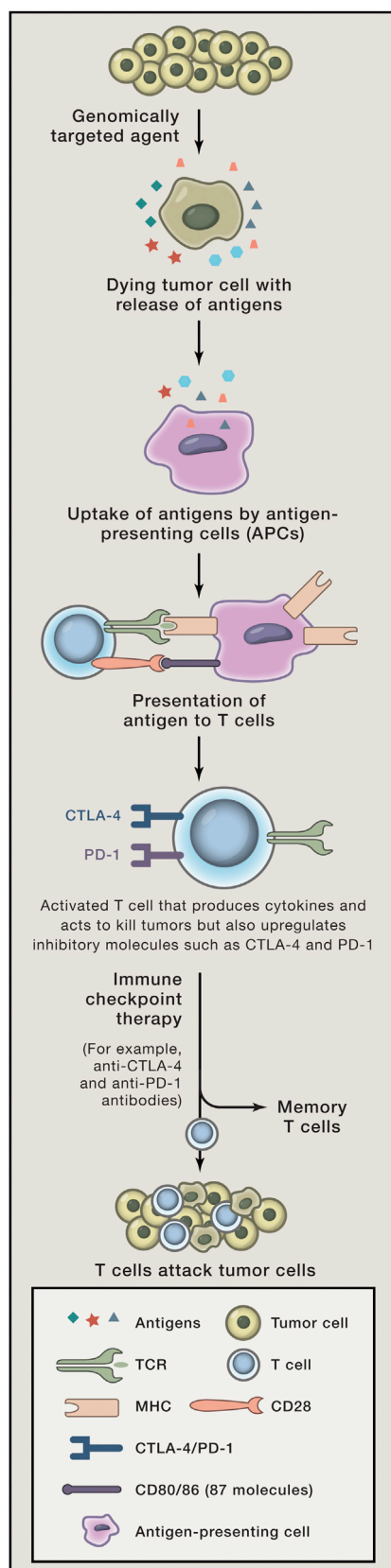
Of these emerging immune checkpoints, LAG-3 is the furthest along in clinical development with a fusion protein (IMP321, Immuntep) and an antibody (BMS-986016, BMS) in clinical trials. The fusion protein was tested as monotherapy in patients with renal cell carcinoma, which was well tolerated and led to stabilization of disease in some patients (Brignone et al., 2009). IMP321 was also tested in combination with paclitaxel chemotherapy in patients with metastatic breast cancer, which led to an objective response rate of 50% (Brignone et al., 2010). Based on these promising results, a phase III clinical trial is expected to begin accrual in 2015. Other clinical trials are ongoing with an antibody against LAG-3 (BMS-986016), which is also being tested in combination with anti-PD-1 (nivolumab) (NCT01968109, <http://www.clinicaltrials.gov>). TIM-3 is another

immune checkpoint for which agents are being developed for clinical testing. Pre-clinical studies indicate that TIM-3 is co-expressed with PD-1 on tumor-infiltrating lymphocytes, and combination therapy targeting these two pathways improves anti-tumor immune responses (Sakuishi et al., 2010). Finally, an antibody targeting VISTA was recently shown to improve anti-tumor immune responses in mice (Le Mercier et al., 2014), with clinical development soon to follow. Again, these agents represent only a partial list of the immune checkpoint agents that are currently under development for clinical testing, with expectations that they will be tested in combination strategies based on in-depth analyses of human tumors to provide an understanding of co-expression of these, and other immunologic targets, to guide rational combinations.

Regarding the co-stimulatory molecules, OX40 and 41BB, which are members of the TNF-receptor superfamily, are furthest along in clinical development. A murine anti-OX40 antibody, given as a single dose, was tested in a phase I clinical trial and found to have an acceptable safety profile, as well as evidence of anti-tumor responses in a subset of patients (Curti et al., 2013). Humanized antibodies against OX40 are expected to enter clinical trial in 2015. Anti-41BB (BMS-663513) is a fully humanized monoclonal antibody that has been tested in a phase I/II study in patients with melanoma, renal cell carcinoma, and ovarian cancer, with promising clinical responses, as well as toxicities, especially at higher doses, which led to re-evaluation of the dose and schedule of treatment (Sznol et al., 2008, J. Clin. Oncol., abstract). Currently, there are five clinical trials with anti-41BB (urelumab, BMS-663513) that are recruiting patients with various tumor types (<http://www.clinicaltrials.gov>), including combination with anti-PD-1 (nivolumab), with data expected to be presented from these trials during the next 1 to 2 years. The third co-stimulatory molecule is inducible co-stimulator (ICOS), a member of the CD28/B7 family whose expression increases on T cells upon T cell activation. ICOS<sup>+</sup> effector T cells (Teff), as opposed to ICOS<sup>+</sup> regulatory T cells (Treg), increase after patients receive treatment with anti-CTLA-4 (Liakou et al., 2008), correlating with clinical benefit in a small retrospective study (Carthon et al., 2010). ICOS thus may serve as a pharmacodynamic biomarker to indicate that anti-CTLA-4 has “hit its target” enhancing T cell activation (Ng Tang et al., 2013). Also, the association of agonistic targeting of ICOS and blockade of CTLA-4 can lead to improved anti-tumor immune responses and tumor rejection in mice (Fan et al., 2014). Anti-ICOS antibodies are expected to enter into clinical trials in 2015. It is likely that combination therapy to simultaneously engage co-stimulatory pathways and limit inhibitory pathways will be a successful path forward to provide clinical benefit. Importantly, based on the profile of toxicities observed to date, it will be critical to closely monitor these combination strategies for potential adjustments of dosage and management of toxicities that may arise.

### Reconciliation: Curative Therapeutic Combinations

The last few decades have witnessed the emergence of two effective but fundamentally different strategies for cancer therapy, each with its own strengths and weaknesses. Genomic-guided identification of mutations that drive cancer has led to

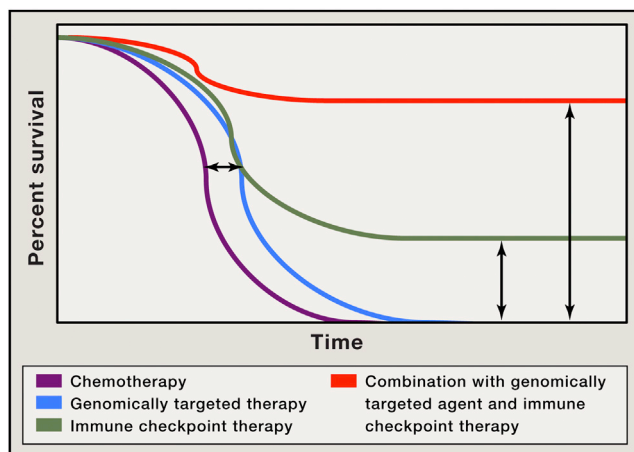


**Figure 1. Combination Therapy May Improve Anti-tumor Responses**

Depiction of tumor cells dying as a result of genomically targeted therapies with release of tumor antigens; tumor antigens are taken up by APCs and are presented in the context of B7 costimulatory molecules to T cells; T cells recognize antigens on APCs to become activated; activated T cells also up-regulate inhibitory checkpoints such as CTLA-4 and PD-1; immune checkpoint therapy prevents attenuation of T cell responses, thereby allowing T cells to kill tumor cells; and T cells may differentiate into memory T cells that can re-activate in the presence of recurrent tumor.

the development of drugs that result in remarkable responses in the majority of patients whose tumors have the targeted lesion, but the responses are relatively short-lived. As was the case with chemotherapies, it is not unreasonable that combinations of genomically targeted agents will be more powerful against cancer than single agents. It is possible that the use of multiple agents may enhance their effectiveness in terms of increasing overall survival. However, the myriad of mechanisms of acquired resistance and the complexity of the target landscape due to inherent genomic instability may prove extremely difficult to overcome through the sole use of genomically targeted strategies, attaining to achieve cure. In contrast, immune checkpoint therapy is inherently multivalent because targeting a single checkpoint can potentially release T cells with specificity for peptides derived from many different antigens present in a tumor, including differentiation, cancer testis, and even neoantigens generated by mutational events inherent in the genomic instability that drives cancer (Snyder et al., 2014; Linnemann et al., 2015). As a result of the generation of improved anti-tumor T cell responses, immune checkpoint therapy results in durable responses but only in a fraction of patients. As discussed in the previous sections, it is certainly possible to target multiple immune checkpoints with different mechanisms for improved anti-tumor responses in greater numbers of patients. Will patients benefit from combination of these two strategies?

Efforts to combine molecularly targeted agents and immunotherapy have already begun. A phase I clinical trial with agents that inhibit receptor tyrosine kinases, sunitinib, or pazopanib, in combination with anti-PD-1, was recently reported and showed promising overall response rates of 40%–50% in patients with metastatic renal cell carcinoma (RCC) (Amin et al., 2014, J. Clin. Oncol., abstract). These types of combinations will require further follow-up to evaluate for survival and durability of responses. An area that has not yet received enough attention is the immunological impact of genetically targeted agents. Vemurafenib, an FDA-approved BRAF inhibitor used for the treatment of melanoma, has been shown to increase expression of tumor antigens and MHC molecules (Frederick et al., 2013), increasing the sensitivity of the tumor cells to immune attack. Vemurafenib also has potent effects on T cells, enhancing the effects of antigen-mediated activation, perhaps as a result of enhanced activation of the MAP kinase pathway after T cell antigen receptor signaling (Atefi et al., 2014). These data suggest that certain agents may be well suited for combination with immunotherapy. However, a clinical trial testing a BRAF inhibitor (vemurafenib) in combination with anti-CTLA-4 (ipilimumab) was terminated due to hepatotoxicity (Ribas et al., 2013). A second clinical trial with a BRAF inhibitor (dabrafenib) in combination with anti-CTLA-4 (ipilimumab) is currently ongoing, and



**Figure 2. Improved Overall Survival as a Result of Combination Therapy**

Depiction of Kaplan-Meier survival curve with genomically targeted agents (blue line) as compared to standard therapies (purple line), indicating an improvement in median overall survival but lack of durable responses; improved median overall survival and durable responses in a fraction of patients treated with immune checkpoint therapy (green line); possibility for improved median overall survival with durable responses for the majority of patients in the setting of combination treatment with genomically targeted agents and immune checkpoint therapy (red line).

preliminary data indicate that this combination appears to be well tolerated (Puzanov et al., 2014, *J. Clin. Oncol.*, abstract), which highlights the need to consider differences in drugs, dose, and/or schedule when evaluating agents for combination strategies. Understanding how different genetically targeted agents affect the responsiveness to immunotherapy may help guide choices of combinations of drugs.

From a mechanistic perspective, it is possible that combination strategies with immune checkpoint therapies and genomically targeted agents will result in induction of immune memory, leading to more durable control of tumor growth than what is achievable with either modality alone. Genomically targeted therapies with high objective response rates actually could serve as “cancer vaccines,” inducing the killing of tumor cells and resulting in the release of tumor antigens and neoantigens, which can then be presented by APCs to tumor-specific T cells (Figure 1). These T cells would become activated but also upregulate inhibitory checkpoints such as CTLA-4 and PD-1, which can be blocked with antibodies to permit enhanced anti-tumor T cell responses, including memory T cell responses, to enable long-term control of disease and possible cure. In addition, the use of targeted agents to directly kill tumor cells, with release of tumor antigens, may focus the activated immune response generated by immunotherapy agents on tumor antigens rather than self-antigens expressed on normal tissues, resulting in fewer adverse events. Furthermore, identification of neoantigens may result in the development of personalized vaccines composed of these neoantigens for novel vaccine strategies plus immune checkpoint agents (Gubin et al., 2014; Tran et al., 2014; Linnemann et al., 2015).

Although it is clear that clinical responses can be elicited with immune checkpoint therapies or genomically targeted agents, it

appears that genomically targeted agents alone tend to improve median survival without providing long-term durable responses (Figure 2, blue line). Targeting immune checkpoints improves median survival but remarkably also provides long-term durable responses, raising the tail of the survival curve (Figure 2, green line). When combined, these therapies are likely to have an additive or even synergistic therapeutic effect that not only would potentially further improve median survival but would also raise the tail of the survival curve, increasing the number of patients that appreciate long-term clinical benefit (Figure 2, red line).

### A Future of Curative Cancer Therapies

Federal funding for research has been overwhelmingly directed toward genomically targeted therapies as compared to immune checkpoint therapies. The fundamental research that led to the identification of CTLA-4 as an immune checkpoint, as well as the pre-clinical studies showing the potential of its blockade in cancer therapy, were funded by the National Cancer Institute, but since then, there have been no major initiatives to accelerate progress in this area. Given the durability of the responses that have been obtained with immune checkpoint therapies, it seems reasonable also to allocate enough funds and resources to research focused on immune checkpoint therapies and combination therapy of genomically targeted agents and immunotherapy with promising curative potential. Efforts to determine the impact of genomically targeted therapies on the immune system should also be prioritized, as they will help to identify which agents can enhance anti-tumor T cell responses and guide the choice of combinations from the two classes of agents. At this stage, it does not seem a stretch to say that increasing funding to combination therapies will be key to development of new safe treatments that may prove to be curative for many patients with many types of cancer.

### ACKNOWLEDGMENTS

P.S. and J.P.A. are founders and advisors for Jounce Therapeutics. P.S. also serves as a consultant for Bristol-Myers Squibb, Amgen, and Glaxo SmithKline. J.P.A. is an inventor of intellectual property owned by the University of California, Berkeley, and licensed to Bristol-Myers Squibb and has received royalties from Bristol-Myers Squibb. Our research is supported by a PCF Challenge Grant in Immunology (J.P.A. and P.S.), NCI/NIH 1-R01 CA1633793-01 (P.S.), Cancer Prevention Research in Texas grants (J.P.A. and P.S.), and a Stand Up To Cancer–Cancer Research Institute Cancer Immunology Dream Team Translational Cancer Research Grant (P.S. and J.P.A.). Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research.

### REFERENCES

- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al. (2013). Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MML-Seq Consortium; ICGC PedBrain (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Ansell, S.M., Lesokhin, A.M., Borrello, I., Halwani, A., Scott, E.C., Gutierrez, M., Schuster, S.J., Millenson, M.M., Cattry, D., Freeman, G.J., et al. (2015). PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N. Engl. J. Med.* 372, 311–319.
- Atefi, M., Avramis, E., Lassen, A., Wong, D.J., Robert, L., Foulad, D., Cerniglia, M., Titz, B., Chodon, T., Graeber, T.G., et al. (2014). Effects of MAPK and PI3K

- pathways on PD-L1 expression in melanoma. *Clin. Cancer Res.* 20, 3446–3457.
- Bishop, J.M. (1991). Molecular themes in oncogenesis. *Cell* 64, 235–248.
- Brahmer, J.R., Tykodi, S.S., Chow, L.Q., Hwu, W.J., Topalian, S.L., Hwu, P., Drake, C.G., Camacho, L.H., Kauh, J., Odunsi, K., et al. (2012). Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N. Engl. J. Med.* 366, 2455–2465.
- Brignone, C., Escudier, B., Grygar, C., Marcu, M., and Triebel, F. (2009). A phase I pharmacokinetic and biological correlative study of IMP321, a novel MHC class II agonist, in patients with advanced renal cell carcinoma. *Clin. Cancer Res.* 15, 6225–6231.
- Brignone, C., Gutierrez, M., Mefti, F., Brain, E., Jarcau, R., Cvitkovic, F., Bousetta, N., Medioni, J., Gligorov, J., Grygar, C., et al. (2010). First-line chemioimmunotherapy in metastatic breast carcinoma: combination of paclitaxel and IMP321 (LAG-3Ig) enhances immune responses and antitumor activity. *J. Transl. Med.* 8, 71.
- Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513, 202–209.
- Carthon, B.C., Wolchok, J.D., Yuan, J., Kamat, A., Ng Tang, D.S., Sun, J., Ku, G., Troncso, P., Logothetis, C.J., Allison, J.P., and Sharma, P. (2010). Preoperative CTLA-4 blockade: tolerability and immune monitoring in the setting of a presurgical clinical trial. *Clin. Cancer Res.* 16, 2861–2871.
- Chambers, C.A., Sullivan, T.J., and Allison, J.P. (1997). Lymphoproliferation in CTLA-4-deficient mice is mediated by costimulation-dependent activation of CD4<sup>+</sup> T cells. *Immunity* 7, 885–895.
- Chapman, P.B., Hauschild, A., Robert, C., Haanen, J.B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., et al.; BRIM-3 Study Group (2011). Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* 364, 2507–2516.
- Chen, Y., McGee, J., Chen, X., Doman, T.N., Gong, X., Zhang, Y., Hamm, N., Ma, X., Higgs, R.E., Bhagwat, S.V., et al. (2014). Identification of druggable cancer driver genes amplified across TCGA datasets. *PLoS ONE* 9, e98293.
- Curran, M.A., Montalvo, W., Yagita, H., and Allison, J.P. (2010). PD-1 and CTLA-4 combination blockade expands infiltrating T cells and reduces regulatory T and myeloid cells within B16 melanoma tumors. *Proc. Natl. Acad. Sci. USA* 107, 4275–4280.
- Curti, B.D., Kovacsics-Bankowski, M., Morris, N., Walker, E., Chisholm, L., Floyd, K., Walker, J., Gonzalez, I., Meeuwse, T., Fox, B.A., et al. (2013). OX40 is a potent immune-stimulating target in late-stage cancer patients. *Cancer Res.* 73, 7189–7198.
- Curtin, J.A., Fridlyand, J., Kageshita, T., Patel, H.N., Busam, K.J., Kutzner, H., Cho, K.H., Aiba, S., Bröcker, E.B., LeBoit, P.E., et al. (2005). Distinct sets of genetic alterations in melanoma. *N. Engl. J. Med.* 353, 2135–2147.
- Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W., et al. (2002). Mutations of the BRAF gene in human cancer. *Nature* 417, 949–954.
- Davis, M.M., and Bjorkman, P.J. (1988). T-cell antigen receptor genes and T-cell recognition. *Nature* 334, 395–402.
- Dong, H., Strome, S.E., Salomao, D.R., Tamura, H., Hirano, F., Flies, D.B., Roche, P.C., Lu, J., Zhu, G., Tamada, K., et al. (2002). Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. *Nat. Med.* 8, 793–800.
- Fan, X., Quezada, S.A., Sepulveda, M.A., Sharma, P., and Allison, J.P. (2014). Engagement of the ICOS pathway markedly enhances efficacy of CTLA-4 blockade in cancer immunotherapy. *J. Exp. Med.* 211, 715–725.
- Frederick, D.T., Piris, A., Cogdill, A.P., Cooper, Z.A., Lezcano, C., Ferrone, C.R., Mitra, D., Boni, A., Newton, L.P., Liu, C., et al. (2013). BRAF inhibition is associated with enhanced melanoma antigen expression and a more favorable tumor microenvironment in patients with metastatic melanoma. *Clin. Cancer Res.* 19, 1225–1231.
- Freeman, G.J., Long, A.J., Iwai, Y., Bourque, K., Chernova, T., Nishimura, H., Fitz, L.J., Malenkovich, N., Okazaki, T., Byrne, M.C., et al. (2000). Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J. Exp. Med.* 192, 1027–1034.
- Gao, J., He, Q., Subudhi, S., Aparicio, A., Zurita-Saavedra, A., Lee, D.H., Jimenez, C., Suarez-Almazor, M., and Sharma, P. (2015). Review of immune-related adverse events in prostate cancer patients treated with ipilimumab: MD Anderson experience. *Oncogene*.
- Greenwald, R.J., Freeman, G.J., and Sharpe, A.H. (2005). The B7 family revisited. *Annu. Rev. Immunol.* 23, 515–548.
- Gros, A., Robbins, P.F., Yao, X., Li, Y.F., Turcotte, S., Tran, E., Wunderlich, J.R., Mixon, A., Farid, S., Dudley, M.E., et al. (2014). PD-1 identifies the patient-specific CD8<sup>+</sup> tumor-reactive repertoire infiltrating human tumors. *J. Clin. Invest.* 124, 2246–2259.
- Gubin, M.M., Zhang, X., Schuster, H., Caron, E., Ward, J.P., Noguchi, T., Ivanova, Y., Hundal, J., Arthur, C.D., Krebber, W.J., et al. (2014). Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* 515, 577–581.
- Hamid, O., Robert, C., Daud, A., Hodi, F.S., Hwu, W.J., Kefford, R., Wolchok, J.D., Hersey, P., Joseph, R.W., Weber, J.S., et al. (2013). Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *N. Engl. J. Med.* 369, 134–144.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Hanna, N., Shepherd, F.A., Fossella, F.V., Pereira, J.R., De Marinis, F., von Pawel, J., Gatzemeier, U., Tsao, T.C., Pless, M., Muller, T., et al. (2004). Randomized phase III trial of pemetrexed versus docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. *J. Clin. Oncol.* 22, 1589–1597.
- Hodi, F.S., Butler, M., Oble, D.A., Seiden, M.V., Haluska, F.G., Kruse, A., Macrae, S., Nelson, M., Canning, C., Lowy, I., et al. (2008). Immunologic and clinical effects of antibody blockade of cytotoxic T lymphocyte-associated antigen 4 in previously vaccinated cancer patients. *Proc. Natl. Acad. Sci. USA* 105, 3005–3010.
- Hodi, F.S., O'Day, S.J., McDermott, D.F., Weber, R.W., Sosman, J.A., Haanen, J.B., Gonzalez, R., Robert, C., Schadendorf, D., Hassel, J.C., et al. (2010). Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* 363, 711–723.
- Hurwitz, A.A., Yu, T.F., Leach, D.R., and Allison, J.P. (1998). CTLA-4 blockade synergizes with tumor-derived granulocyte-macrophage colony-stimulating factor for treatment of an experimental mammary carcinoma. *Proc. Natl. Acad. Sci. USA* 95, 10067–10071.
- Ishida, Y., Agata, Y., Shibahara, K., and Honjo, T. (1992). Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J.* 11, 3887–3895.
- Jenkins, M.K., and Schwartz, R.H. (1987). Antigen presentation by chemically modified splenocytes induces antigen-specific T cell unresponsiveness in vitro and in vivo. *J. Exp. Med.* 165, 302–319.
- Keir, M.E., Butte, M.J., Freeman, G.J., and Sharpe, A.H. (2008). PD-1 and its ligands in tolerance and immunity. *Annu. Rev. Immunol.* 26, 677–704.
- Knudson, A.G. (2001). Two genetic hits (more or less) to cancer. *Nat. Rev. Cancer* 1, 157–162.
- Krummel, M.F., and Allison, J.P. (1995). CD28 and CTLA-4 have opposing effects on the response of T cells to stimulation. *J. Exp. Med.* 182, 459–465.
- Kwak, E.L., Bang, Y.J., Camidge, D.R., Shaw, A.T., Solomon, B., Maki, R.G., Ou, S.H., Dezube, B.J., Jänne, P.A., Costa, D.B., et al. (2010). Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* 363, 1693–1703.
- Kwon, E.D., Drake, C.G., Scher, H.I., Fizazi, K., Bossi, A., van den Eertwegh, A.J., Krainer, M., Houede, N., Santos, R., Mahammed, H., et al.; CA184-043 Investigators (2014). Ipilimumab versus placebo after radiotherapy in patients with metastatic castration-resistant prostate cancer that had progressed after docetaxel chemotherapy (CA184-043): a multicentre, randomised, double-blind, phase 3 trial. *Lancet Oncol.* 15, 700–712.

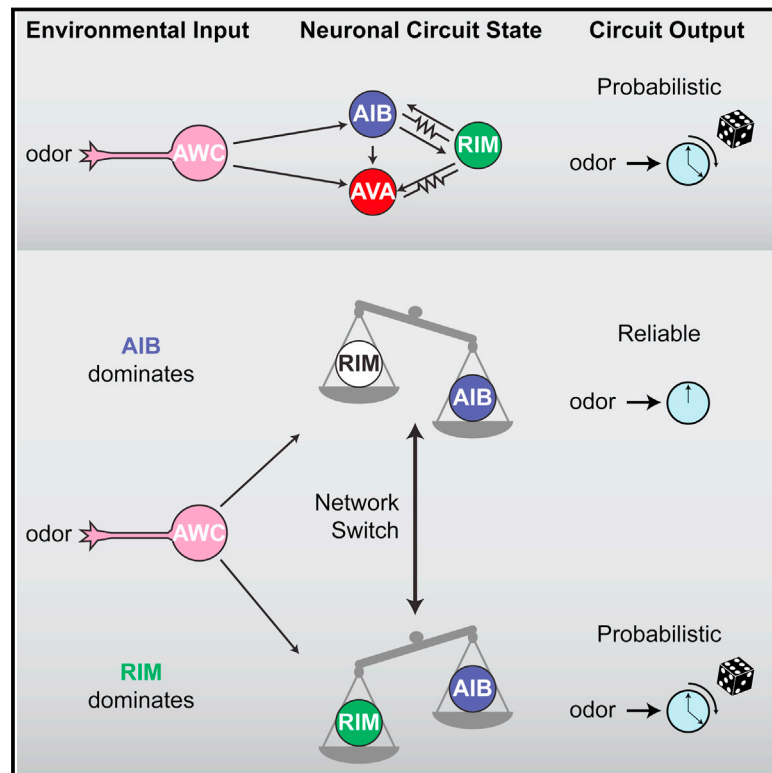


- Larkin, J., Ascierto, P.A., Dréno, B., Atkinson, V., Liszkay, G., Maio, M., Mandalà, M., Demidov, L., Stroyakovskiy, D., Thomas, L., et al. (2014). Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *N. Engl. J. Med.* 371, 1867–1876.
- Le Mercier, I., Chen, W., Lines, J.L., Day, M., Li, J., Sergent, P., Noelle, R.J., and Wang, L. (2014). VISTA regulates the development of protective antitumor immunity. *Cancer Res.* 74, 1933–1944.
- Leach, D.R., Krummel, M.F., and Allison, J.P. (1996). Enhancement of anti-tumor immunity by CTLA-4 blockade. *Science* 271, 1734–1736.
- Liakou, C.I., Kamat, A., Tang, D.N., Chen, H., Sun, J., Troncso, P., Logothetis, C., and Sharma, P. (2008). CTLA-4 blockade increases IFN $\gamma$ -producing CD4<sup>+</sup>ICOS<sup>+</sup> cells to shift the ratio of effector to regulatory T cells in cancer patients. *Proc. Natl. Acad. Sci. USA* 105, 14987–14992.
- Linnemann, C., van Buuren, M.M., Bies, L., Verdegaal, E.M., Schotte, R., Calis, J.J., Behjati, S., Velds, A., Hilkman, H., Atmioui, D.E., et al. (2015). High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4<sup>+</sup> T cells in human melanoma. *Nat. Med.* 21, 81–85.
- Martz, C.A., Ottina, K.A., Singleton, K.R., Jasper, J.S., Wardell, S.E., Peraz-Penton, A., Anderson, G.R., Winter, P.S., Wang, T., Alley, H.M., et al. (2014). Systematic identification of signaling pathways with potential to confer anti-cancer drug resistance. *Sci. Signal.* 7, ra121.
- Melero, I., Shuford, W.W., Newby, S.A., Aruffo, A., Ledbetter, J.A., Hellström, K.E., Mittler, R.S., and Chen, L. (1997). Monoclonal antibodies against the 4-1BB T-cell activation molecule eradicate established tumors. *Nat. Med.* 3, 682–685.
- Melero, I., Gaudernack, G., Gerritsen, W., Huber, C., Parmiani, G., Scholl, S., Thatcher, N., Wagstaff, J., Zielinski, C., Faulkner, I., and Mellstedt, H. (2014). Therapeutic vaccines for cancer: an overview of clinical trials. *Nat. Rev. Clin. Oncol.* 11, 509–524.
- Mueller, D.L., Jenkins, M.K., and Schwartz, R.H. (1989). An accessory cell-derived costimulatory signal acts independently of protein kinase C activation to allow T cell proliferation and prevent the induction of unresponsiveness. *J. Immunol.* 142, 2617–2628.
- Ng Tang, D., Shen, Y., Sun, J., Wen, S., Wolchok, J.D., Yuan, J., Allison, J.P., and Sharma, P. (2013). Increased frequency of ICOS<sup>+</sup> CD4<sup>+</sup> T cells as a pharmacodynamic biomarker for anti-CTLA-4 therapy. *Cancer Immunol. Res.* 1, 229–234.
- Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer.* 12, 252–264.
- Peggs, K.S., Quezada, S.A., Chambers, C.A., Korman, A.J., and Allison, J.P. (2009). Blockade of CTLA-4 on both effector and regulatory T cell compartments contributes to the antitumor activity of anti-CTLA-4 antibodies. *J. Exp. Med.* 206, 1717–1725.
- Postow, M.A., Callahan, M.K., and Wolchok, J.D. (2015). Immune checkpoint blockade in cancer therapy. *J. Clin. Oncol.* Published online January 20, 2015. <http://dx.doi.org/10.1200/JCO.2014.59.4358>.
- Powles, T., Eder, J.P., Fine, G.D., Braiteh, F.S., Lort, Y., Cruz, C., Bellmunt, J., Burris, H.A., Petrylak, D.P., Teng, S.L., et al. (2014). MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer. *Nature* 515, 558–562.
- Quezada, S.A., Peggs, K.S., Curran, M.A., and Allison, J.P. (2006). CTLA4 blockade and GM-CSF combination immunotherapy alters the intratumor balance of effector and regulatory T cells. *J. Clin. Invest.* 116, 1935–1945.
- Ribas, A., Hodi, F.S., Callahan, M., Konto, C., and Wolchok, J. (2013). Hepatotoxicity with combination of vemurafenib and ipilimumab. *N. Engl. J. Med.* 368, 1365–1366.
- Robert, C., Thomas, L., Bondarenko, I., O'Day, S., Weber, J., Garbe, C., Lebbe, C., Baurain, J.F., Testori, A., Grob, J.J., et al. (2011). Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. *N. Engl. J. Med.* 364, 2517–2526.
- Robert, C., Karaszewska, B., Schachter, J., Rutkowski, P., Mackiewicz, A., Stroiakovski, D., Lichinitser, M., Dummer, R., Grange, F., Mortier, L., et al. (2015a). Improved overall survival in melanoma with combined dabrafenib and trametinib. *N. Engl. J. Med.* 372, 30–39.
- Robert, C., Long, G.V., Brady, B., Dutriaux, C., Maio, M., Mortier, L., Hassel, J.C., Rutkowski, P., McNeil, C., Kalinka-Warzocha, E., et al. (2015b). Nivolumab in previously untreated melanoma without BRAF mutation. *N. Engl. J. Med.* 372, 320–330.
- Rosenberg, S.A., Yang, J.C., and Restifo, N.P. (2004). Cancer immunotherapy: moving beyond current vaccines. *Nat. Med.* 10, 909–915.
- Sakuishi, K., Apetoh, L., Sullivan, J.M., Blazar, B.R., Kuchroo, V.K., and Anderson, A.C. (2010). Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *J. Exp. Med.* 207, 2187–2194.
- Schadendorf, D., Hodi, F.S., Robert, C., Weber, J.S., Margolin, K., Hamid, O., Patt, D., Chen, T.T., Berman, D.M., and Wolchok, J.D. (2015). Pooled analysis of long-term survival data from phase II and phase III trials of ipilimumab in unresectable or metastatic melanoma. *J. Clin. Oncol.* Published online February 9, 2015. <http://dx.doi.org/10.1200/JCO.2014.56.2736>.
- Segal, N.H., Parsons, D.W., Peggs, K.S., Velculescu, V., Kinzler, K.W., Vogelstein, B., and Allison, J.P. (2008). Epitope landscape in breast and colorectal cancer. *Cancer Res.* 68, 889–892.
- Sharma, P., Wagner, K., Wolchok, J.D., and Allison, J.P. (2011). Novel cancer immunotherapy agents with survival benefit: recent successes and next steps. *Nat. Rev. Cancer* 11, 805–812.
- Shaw, A.T., Kim, D.W., Nakagawa, K., Seto, T., Crinó, L., Ahn, M.J., De Pas, T., Besse, B., Solomon, B.J., Blackhall, F., et al. (2013). Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N. Engl. J. Med.* 368, 2385–2394.
- Sjöblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274.
- Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J.M., Desrichard, A., Walsh, L.A., Postow, M.A., Wong, P., Ho, T.S., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* 371, 2189–2199.
- Soda, M., Choi, Y.L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H., et al. (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448, 561–566.
- Solomon, E., Borrow, J., and Goddard, A.D. (1991). Chromosome aberrations and cancer. *Science* 254, 1153–1160.
- Sosman, J.A., Kim, K.B., Schuchter, L., Gonzalez, R., Pavlick, A.C., Weber, J.S., McArthur, G.A., Hutson, T.E., Moschos, S.J., Flaherty, K.T., et al. (2012). Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib. *N. Engl. J. Med.* 366, 707–714.
- Tang, C., Wang, X., Soh, H., Seyedin, S., Cortez, M.A., Krishnan, S., Massarelli, E., Hong, D., Naing, A., Diab, A., et al. (2014). Combining radiation and immunotherapy: a new systemic therapy for solid tumors? *Cancer Immunol. Res.* 2, 831–838.
- Tivol, E.A., Borriello, F., Schweitzer, A.N., Lynch, W.P., Bluestone, J.A., and Sharpe, A.H. (1995). Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity* 3, 541–547.
- Topalian, S.L., Hodi, F.S., Brahmer, J.R., Gettinger, S.N., Smith, D.C., McDermott, D.F., Powderly, J.D., Carvajal, R.D., Sosman, J.A., Atkins, M.B., et al. (2012). Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* 366, 2443–2454.
- Townsend, S.E., and Allison, J.P. (1993). Tumor rejection after direct costimulation of CD8<sup>+</sup> T cells by B7-transfected melanoma cells. *Science* 259, 368–370.
- Tran, E., Turcotte, S., Gros, A., Robbins, P.F., Lu, Y.C., Dudley, M.E., Wunderlich, J.R., Somerville, R.P., Hogan, K., Hinrichs, C.S., et al. (2014). Cancer immunotherapy based on mutation-specific CD4<sup>+</sup> T cells in a patient with epithelial cancer. *Science* 344, 641–645.

- Triebel, F., Jitsukawa, S., Baixeras, E., Roman-Roman, S., Genevee, C., Viegas-Pequignot, E., and Hercend, T. (1990). LAG-3, a novel lymphocyte activation gene closely related to CD4. *J. Exp. Med.* *171*, 1393–1405.
- van den Eertwegh, A.J., Versluis, J., van den Berg, H.P., Santegoets, S.J., van Moorselaar, R.J., van der Sluis, T.M., Gall, H.E., Harding, T.C., Jooss, K., Lowy, I., et al. (2012). Combined immunotherapy with granulocyte-macrophage colony-stimulating factor-transduced allogeneic prostate cancer cells and ipilimumab in patients with metastatic castration-resistant prostate cancer: a phase 1 dose-escalation trial. *Lancet Oncol.* *13*, 509–517.
- van Elsas, A., Hurwitz, A.A., and Allison, J.P. (1999). Combination immunotherapy of B16 melanoma using anti-cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) and granulocyte/macrophage colony-stimulating factor (GM-CSF)-producing vaccines induces rejection of subcutaneous and metastatic tumors accompanied by autoimmune depigmentation. *J. Exp. Med.* *190*, 355–366.
- Waitz, R., Fassò, M., and Allison, J.P. (2012). CTLA-4 blockade synergizes with cryoablation to mediate tumor rejection. *Onc Immunology* *1*, 544–546.
- Walunas, T.L., Lenschow, D.J., Bakker, C.Y., Linsley, P.S., Freeman, G.J., Green, J.M., Thompson, C.B., and Bluestone, J.A. (1994). CTLA-4 can function as a negative regulator of T cell activation. *Immunity* *1*, 405–413.
- Wang, L., Rubinstein, R., Lines, J.L., Wasiuk, A., Ahonen, C., Guo, Y., Lu, L.F., Gondek, D., Wang, Y., Fava, R.A., et al. (2011). VISTA, a novel mouse Ig superfamily ligand that negatively regulates T cell responses. *J. Exp. Med.* *208*, 577–592.
- Waterhouse, P., Penninger, J.M., Timms, E., Wakeham, A., Shahinian, A., Lee, K.P., Thompson, C.B., Griesser, H., and Mak, T.W. (1995). Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science* *270*, 985–988.
- Weinberg, R.A. (1991). Tumor suppressor genes. *Science* *254*, 1138–1146.
- Wolchok, J.D., Kluger, H., Callahan, M.K., Postow, M.A., Rizvi, N.A., Lesokhin, A.M., Segal, N.H., Ariyan, C.E., Gordon, R.A., Reed, K., et al. (2013). Nivolumab plus ipilimumab in advanced melanoma. *N. Engl. J. Med.* *369*, 122–133.
- Yang, J.C., Hughes, M., Kammula, U., Royal, R., Sherry, R.M., Topalian, S.L., Suri, K.B., Levy, C., Allen, T., Mavroukakis, S., et al. (2007). Ipilimumab (anti-CTLA4 antibody) causes regression of metastatic renal cell cancer associated with enteritis and hypophysitis. *J. Immunother.* *30*, 825–830.
- Zamarin, D., Holmgard, R.B., Subudhi, S.K., Park, J.S., Mansour, M., Palese, P., Merghoub, T., Wolchok, J.D., and Allison, J.P. (2014). Localized oncolytic virotherapy overcomes systemic tumor resistance to immune checkpoint blockade immunotherapy. *Sci. Transl. Med.* *6*, 226ra32.

# Feedback from Network States Generates Variability in a Probabilistic Olfactory Circuit

## Graphical Abstract



## Authors

Andrew Gordus, Navin Pokala, ..., Steven W. Flavell, Cornelia I. Bargmann

## Correspondence

cori@rockefeller.edu

## In Brief

Even when a stimulus invariably activates a sensory neuron, the motor output/behavior is probabilistic because of variability in the network state of interneurons downstream of the sensory neuron. Manipulating the activity of these interneurons relative to one another can drive olfactory responses in *C. elegans* to be more deterministic.

## Highlights

- Interneurons in an olfactory circuit have variable responses to a fixed odor input
- Interneurons participate in collective network states that correlate with behavior
- Reliability of the AIB interneuron's odor response depends on the network activity state
- Chemical synapses from the RIM interneuron increase variability of the odor response



# Feedback from Network States Generates Variability in a Probabilistic Olfactory Circuit

Andrew Gordus,<sup>1</sup> Navin Pokala,<sup>1</sup> Sagi Levy,<sup>1</sup> Steven W. Flavell,<sup>1</sup> and Cornelia I. Bargmann<sup>1,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute and Lulu and Anthony Wang Laboratory of Neural Circuits and Behavior, The Rockefeller University, New York, NY 10065, USA

\*Correspondence: [cori@rockefeller.edu](mailto:cori@rockefeller.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.018>

## SUMMARY

Variability is a prominent feature of behavior and is an active element of certain behavioral strategies. To understand how neuronal circuits control variability, we examined the propagation of sensory information in a chemotaxis circuit of *C. elegans* where discrete sensory inputs can drive a probabilistic behavioral response. Olfactory neurons respond to odor stimuli with rapid and reliable changes in activity, but downstream AIB interneurons respond with a probabilistic delay. The interneuron response to odor depends on the collective activity of multiple neurons—AIB, RIM, and AVA—when the odor stimulus arrives. Certain activity states of the network correlate with reliable responses to odor stimuli. Artificially generating these activity states by modifying neuronal activity increases the reliability of odor responses in interneurons and the reliability of the behavioral response to odor. The integration of sensory information with network states may represent a general mechanism for generating variability in behavior.

## INTRODUCTION

Variability is intrinsic to behavior. The behavioral response of individuals to a defined sensory stimulus varies from trial to trial, even when it is predictable on average. Although variability may limit task performance, both behavioral and theoretical analyses suggest that it can also be a creative element of behavioral strategies (Thrun, 1992; Hessler and Doupe, 1999; Ölveczky et al., 2005; Tumer and Brainard, 2007; Chaisanguanthum et al., 2014). In foraging animals, behavioral variation over short and long timescales allows efficient exploration of environments with unevenly distributed resources (Charnov, 1976; Humphries et al., 2010). In an analogous fashion, computer machine-learning algorithms use variability to escape local minima and reach global optima (Kirkpatrick et al., 1983; Mitsutake et al., 2013). Game-theoretical approaches suggest that variable strategies are often the best responses to unpredictable conditions, particularly in the presence of competitors or predictors (Harsanyi, 1973). At a neuronal level, intrinsically generated variability provides a substrate for reward learning, and increased variability has been linked to enhanced learning in motor tasks

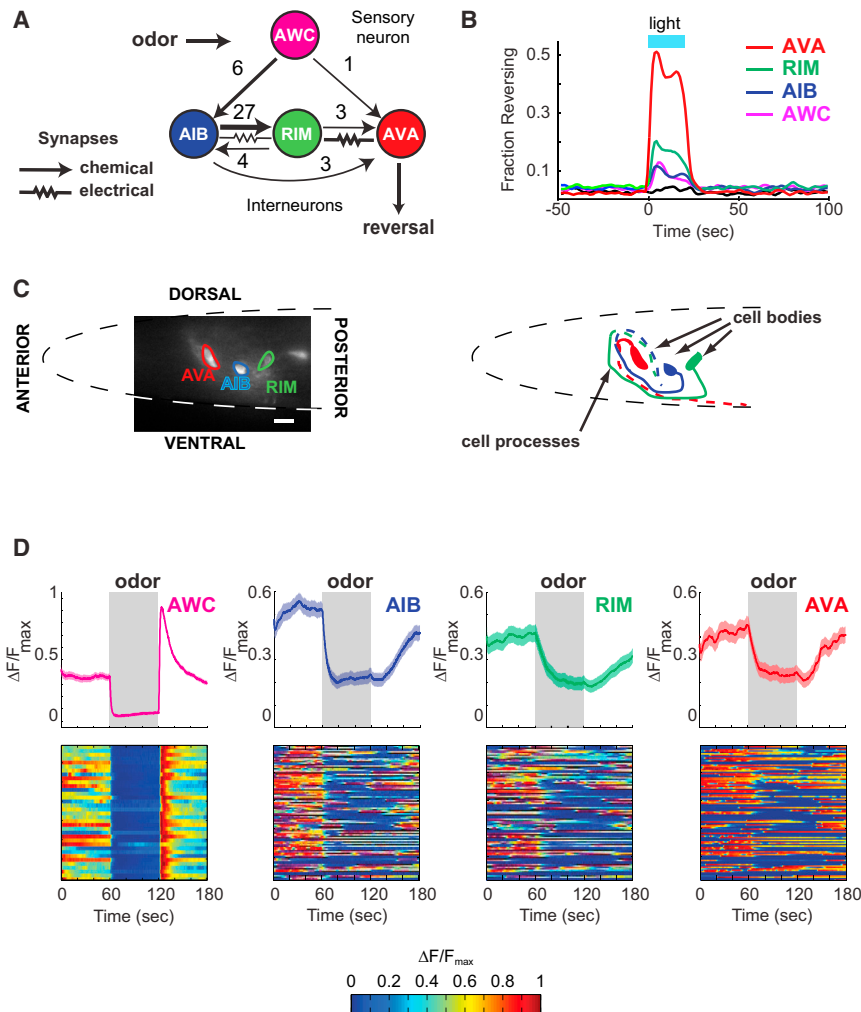
(Ölveczky et al., 2005; Tumer and Brainard, 2007; Chaisanguanthum et al., 2014).

Trial-to-trial variability in responses to a sensory stimulus can result from several mechanisms. There is unavoidable noise in sensory systems operating near their detection or discrimination thresholds (Barlow et al., 1971; Lillywhite and Laughlin, 1979; Bialek, 1987). This stochastic noise decreases precision, but it can enhance sensitivity to weak signals (Benzi et al., 1981; Longtin et al., 1991). At subsequent levels, noise in synaptic transmission or cellular properties can alter signal propagation at any point between sensory and motor systems. Finally, the state of the neuronal network when a signal arrives can influence the network response, especially if its dynamics are highly sensitive to initial conditions (Rajan et al., 2010). However, it is challenging to ascribe single-trial variation to a precise source in complex systems in which the neuronal source of behavioral variation must be indirectly inferred from population measurements of neuronal activity.

The compact nervous system of the nematode worm *Caenorhabditis elegans*, which has only 302 neurons and about 7,000 connections (White et al., 1986), provides an opportunity to address the neuronal sources of behavioral variability. Variability is an explicit element of *C. elegans* behavioral strategies for locating attractants. As first described in bacteria, a biased random walk allows organisms to approach an attractant source by changing their turning rates on the basis of whether stimulus concentrations are increasing or decreasing (Berg and Brown, 1972). In this probabilistic behavior, the rate of turning is predictable, but individual reorientation events are not. *C. elegans* has probabilistic reversal (reorientation) responses to odors, tastes, and temperature associated with chemotaxis and thermotaxis behaviors (Pierce-Shimomura et al., 1999; Clark et al., 2007). The sensory neurons and circuits for these behaviors have been extensively characterized, but it is not known where in the circuit a decision is made to reorient movement.

*C. elegans* neurons fall into three computational levels: sensory neurons that gather information, motor neurons that synapse onto muscle, and extensively interconnected interneurons. *C. elegans* chemotaxis to attractive odors such as isoamyl alcohol (IAA) is initiated by two AWC olfactory neurons. Attractive odors decrease AWC calcium levels and suppress reversal behaviors as part of a biased random walk strategy, whereas odor removal increases AWC calcium and stimulates reversals (Chalasani et al., 2007; Albrecht and Bargmann, 2011). The AWC calcium response, which is likely correlated





**Figure 1. Calcium Dynamics of AIB, RIM, and AVA Neurons in Response to Odor**

(A) Simplified wiring diagram showing AWC sensory neurons and three interneurons in a circuit linking AWC to reversal behavior, and the number of direct synapses between each pair of neurons (White et al., 1986). A more complete circuit appears in Figure 7.

(B) Light-induced reversal behaviors in wild-type animals expressing Channelrhodopsin2 (ChR2) in specific neurons, showing the instantaneous fraction of animals reversing; the total percentage of animals that respond is higher (Figure S1A). The low light levels used here (0.025 mW/mm<sup>2</sup>) did not activate the endogenous *C. elegans* light avoidance response (black line).

(C) Single frame showing AIB, RIM, and AVA neurons expressing GCaMP3 in an animal restrained in the microfluidic imaging chip, and schematic showing the head of the animal and location of neurons and processes. Dashed lines represent processes on the contralateral side of the head; the AIB, RIM, and AVA neurons from the contralateral side are not shown. Scale bar, 10  $\mu$ m.

(D) Averaged calcium responses to a 1-min exposure to 92  $\mu$ M IAA (top) and individual traces (bottom). AIB, RIM, and AVA were recorded simultaneously and aligned in the same order in each panel ( $n = 83$ ); AWC was recorded independently ( $n = 35$ ). Calcium dynamics were normalized to peak and trough values for each trace before averaging. Shaded area is SEM.

with depolarization, is highly reliable from trial to trial, even after dozens of odor presentations (Larsch et al., 2013). By contrast, the reversal response is probabilistic. Even under well-controlled conditions, animals may or may not reverse on individual trials, regardless of the strength of the AWC calcium response (Larsch et al., 2013). Most reversals rely upon the two AVA command interneurons, which synapse onto motor neurons that control the final common pathway for transitions from forward to backward movement (Chalfie et al., 1985). An increase in AVA calcium activity consistently correlates with the beginning of a reversal, and a decrease with its termination (Chronis et al., 2007; Ben Arous et al., 2010; Faumont et al., 2011). Thus, the variability in the behavioral response results from variable transmission of information from the AWC sensory neuron to AVA command neurons.

The *C. elegans* connectome provides a framework to examine the intermediate steps of information propagation at the single-neuron level. Here we show that the AIB and RIM interneurons that link AWC to AVA integrate sensory information with ongoing network states to produce neuronal and behavioral variability. We find that the instantaneous activity state of the integrating

network can predict its response to odor stimuli and that artificially generating certain activity states by modifying RIM activity can increase the reliability of AVA responses and behavior. Internal network states may serve as a source of variability that influences the neuronal response to sensory input and ultimately the animal's response to its environment.

## RESULTS

### Odor-Evoked Calcium Responses in AIB, RIM, and AVA

Among several neuronal pathways connecting AWC to motor output, four pairs of neurons—AWC, AIB, RIM, and AVA—represent a starting point for defining a connectivity diagram for reversals (White et al., 1986; Figure 1A). The *C. elegans* wiring diagram predicts a single direct synaptic connection between AWC sensory neurons and AVA backward command neurons, but a much greater number of indirect connections. Most reversals initiated by AWC require the two AIB interneurons, which are major synaptic targets of AWC and many other sensory neurons (Gray et al., 2005; White et al., 1986). AIB has a few direct synapses onto AVA and a much stronger indirect connection to AVA through the two RIM interneurons, which are connected to both AIB and AVA by chemical and electrical synapses. RIM neurons also form neuromuscular junctions that affect head movements, which were not studied here.

AWC, AIB, and AVA stimulate odor-evoked and spontaneous reversals (Gray et al., 2005; Chalasani et al., 2007; Guo et al., 2009), but experiments conducted in different conditions and genetic backgrounds have led to ambiguous conclusions about whether RIM stimulates or inhibits reversals (Gray et al., 2005; Guo et al., 2009; Piggott et al., 2011). To clarify this relationship, we depolarized each neuron type individually in wild-type animals by cell-specific expression and activation of Channelrhodopsin2 (Nagel et al., 2005). Acute light stimulation of AWC, AIB, RIM, or AVA resulted in increased reversal behaviors (Figure 1B; Figure S1), suggesting that the net activity of each neuron promotes reversals.

To monitor information flow between these neurons, we imaged calcium in animals expressing the genetically encoded calcium indicator GCaMP3 (Tian et al., 2009) in AWC, AIB, RIM, and AVA individually and in combinations. Animals were restrained in a small microfluidic chamber that allowed precise delivery and removal of odor stimuli to the nose (Figures 1C and S1) (Chronis et al., 2007). AWC and AIB have previously been examined in this imaging system (Chalasani et al., 2007), but RIM and AVA have not. AWC responds to odor addition with an immediate reduction in calcium and to odor removal with a sharp calcium rise followed by a return to baseline levels (Figure 1D). AIB, RIM, and AVA responded to odor addition with a calcium decrease relative to the average baseline, and to odor removal with a slow return toward the baseline, without an overshoot (Figure 1D). The interneuron calcium responses appeared smaller in magnitude and slower than those of AWC neurons. We focused subsequent analysis on odor addition, in part because of its robustness and in part because the more complex response to odor removal is regulated by odor history as well as concentration (Chalasani et al., 2007).

### AIB, RIM, and AVA Have Distinct High and Low Activity States

Examination of individual traces revealed an unexpected feature of calcium signals in AIB, RIM, and AVA not visualized in the averages: the neurons appeared to switch between long-lasting high and low calcium states, spending little time at intermediate values (Figures 1D and 2A–2C). Furthermore, transitions between high and low states for all three neurons often occurred at the same time, both spontaneously in buffer (Figure 2A) and in response to odor stimuli (Figure 2C). Quantitative analysis confirmed that AIB, RIM, and AVA calcium signals had a bimodal distribution, with a strong bias toward distinct high and low states (Figure 2B). While *C. elegans* neurons lack classical sodium-based action potentials, they do have voltage-activated channels that can generate active properties such as bistability (Goodman et al., 1998; Mellem et al., 2008).

Calcium signals in *C. elegans* neurons are generally correlated with depolarization, but can vary between cellular compartments (Chalasani et al., 2007; Hendricks et al., 2012). The presynaptic calcium that drives neurotransmitter release is most relevant to neuronal function and can be observed directly by monitoring GCaMP signals in axons. Like somatic responses, calcium signals in the axons of AIB, RIM, and AVA neurons were bimodal, with long-lasting high and low states (Figure S2A). They began to rise or fall at the same time as somatic calcium signals, but

the response magnitude peaked more quickly in the axon, especially for RIM (Figure 2D). Odor-evoked activity in axons rose and fell with similar dynamics in AIB, RIM, and AVA (Figure 2D). Because of the proximity of AIB, RIM, and AVA axons, simultaneous imaging was only possible for cell bodies.

We defined distinct ON and OFF states for AIB, RIM, and AVA on the basis of the beginning of the rise or fall in activity, which was synchronous in cell bodies and axons (Figure S2B). Both ON and OFF states varied greatly in duration, with lengths that ranged from a few seconds to several minutes in animals held in constant conditions (Figure S2C).

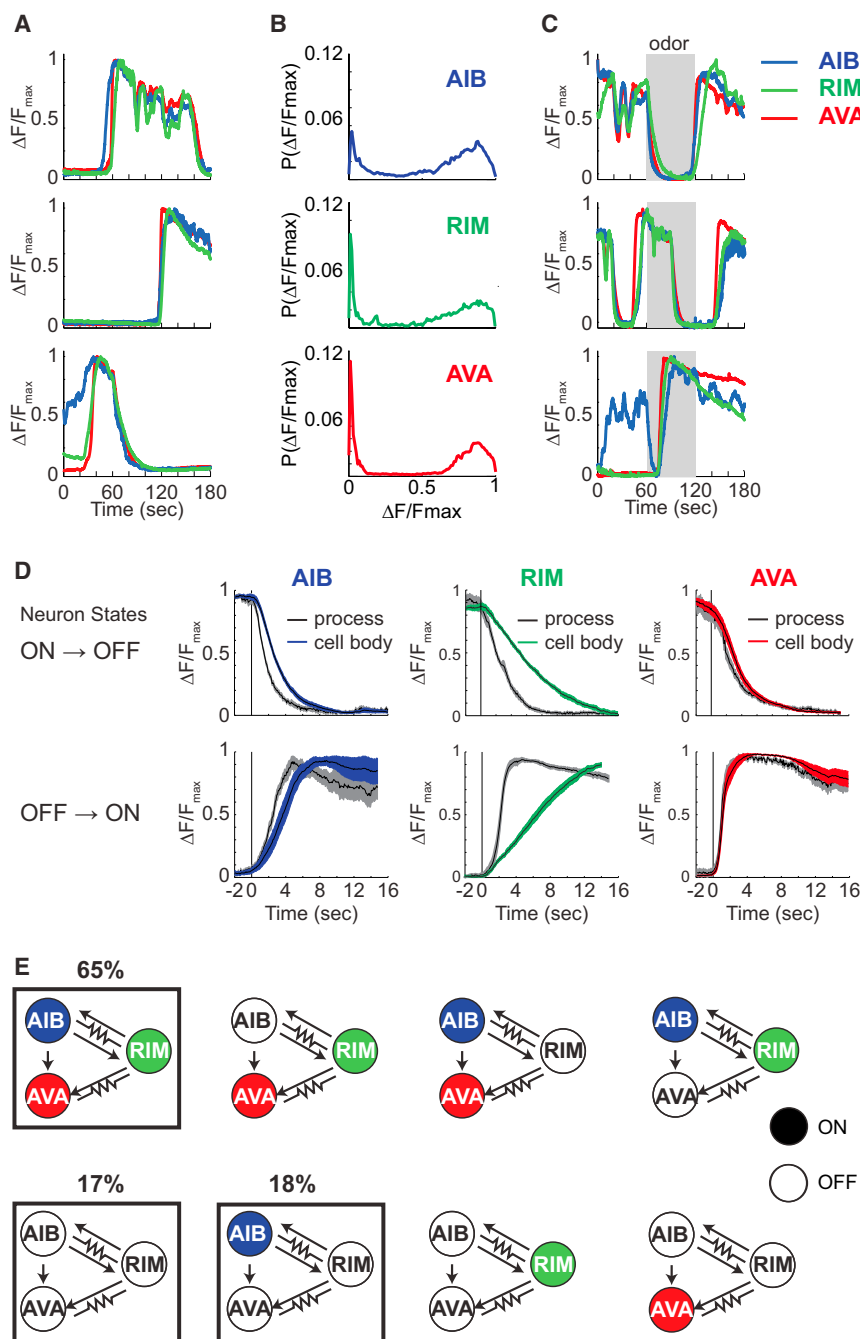
Correlated activity among a set of neurons that includes AIB and AVA has also been observed in whole-brain imaging (Schröder et al., 2013; Prevedel et al., 2014), in agreement with these observations. To assess the degree of correlation between the interneurons, AIB, RIM, and AVA were simultaneously imaged and then independently classified into ON or OFF states for animals imaged for 1 min in buffer (Figures S2 and S3). Most neuronal state transitions were correlated to within a few seconds, which is at the limit of resolution of the binary classification scheme. AVA produced the clearest transitions between ON and OFF states, whereas the slower calcium dynamics of RIM led to less precise transition assignments. AIB had relatively fast calcium dynamics and also displayed more low-amplitude high frequency (second, subsecond) dynamics in the ON state that made shorter ON/OFF assignments less precise (Figures 2A and 2C; Figure S3).

Of the eight possible binary states of AIB, RIM, and AVA, only three occurred for time frames longer than 10 s: (1) all neurons ON, (2) all neurons OFF, or (3) AIB only (AIB-ON, AVA/RIM-OFF) (Figure 2E; Figure S2D). The AIB-only state could last for many seconds when it did occur (Figure S2D). Mutual information analysis confirmed that AVA and RIM are more tightly coupled to one another than to AIB (Experimental Procedures; Figure S3). Transitions between the three network states continued over several hours in animals that were physically immobilized in the absence of externally applied sensory stimuli or pharmacological agents.

### Odor Addition Drives State Transitions in AIB, RIM, and AVA Interneurons

IAA addition induces a rapid and reliable suppression of activity in AWC sensory neurons (Figure 1D; Chalasani et al., 2007). Odor effects on interneurons were more variable and were most easily understood in the context of the distinct ON and OFF network states.

We first considered each interneuron independently. The average intermediate response to odor addition (Figure 1D) resulted from some neurons that responded very strongly soon after odor addition and others that did not respond at all. This effect was most evident when neuronal activity profiles were binned according to their state just prior to odor addition and sorted on the basis of the time to the next transition (Figure 3A). A majority of the AIB, RIM, and AVA neurons in the ON state responded to odor with a transition to the OFF state within a few seconds, and a minority did not. Neurons in the OFF state remained OFF after odor addition, presumably because they could not be suppressed below this apparent baseline. To quantify the



**Figure 2. AIB, RIM, and AVA Have Bistable, Correlated Activity States**

(A) Representative calcium dynamics in simultaneously recorded AIB, RIM, and AVA neurons from three animals in buffer.

(B) Probability distribution of normalized calcium activity showing bimodal distribution of activity in AIB, RIM, and AVA neurons in buffer ( $n = 83$ ).

(C) Representative calcium dynamics in simultaneously recorded AIB, RIM, and AVA neurons from three animals in response to a 1-min odor pulse (gray bars).

(D) Calcium dynamics in simultaneously imaged cell bodies (color) and processes (gray), aligned to ON→OFF and OFF→ON transitions ( $n = 17$ ). Shaded areas are SEM. Calcium dynamics from individually imaged neurons are normalized to peak and trough values for each trace in all panels.

(E) The eight possible network states and the observed frequencies of each state lasting longer than 10 s during a 1-min period prior to odor exposure ( $n = 83$ ).

ure S3E). These results suggest that odor drives collective all-or-none transitions in multiple interneurons to shift network states.

Odor addition had other effects on interneuron activity as well (Figures 3D–3H). For all three interneurons, the median duration of the initial OFF response induced by odor was 5- to 10-fold longer than in buffer controls (Figure 3E; Figure S2C), and the total fraction of time spent in the ON state decreased 2- to 4-fold (Figure 3F). In addition to large-scale ON-OFF transitions, odor often elicited a rapid, small-amplitude decrease in the calcium signal of the AIB neuron, which was not always accompanied by a full OFF state or by similar changes in the RIM and AVA neurons (Figures 3G and 3H; Figures S3C and S3D).

The odor-regulated, inefficient, all-or-none transitions in neuronal activity of AIB, RIM, and AVA seemed of particular interest for probabilistic reversal behavior: they captured both the delay and the variability characteristic of the behavioral

effect of odor addition, we measured the time after odor exposure at which each interneuron switched from an ON to OFF state and compared with controls switched from buffer to buffer (Figures 3B–3D). Odor shifted the entire distribution of AIB, RIM, and AVA neurons toward OFF states within a few seconds, an effect that was strongest for AIB.

The responses of simultaneously imaged AIB, RIM, and AVA neurons to odor addition were highly correlated with one another; in most cases, either all three neurons shifted from an ON to an OFF state at similar times or none did (Figure 1D; Fig-

output (Larsch et al., 2013). These transitions transform reliable AWC activity into variable AVA responses.

### RIM Neurons Create Variability in AIB Odor Responses

An insight into the source of interneuron variability was provided by the network state in which only AIB was ON. In AIB-only states, unlike all-ON states, odor addition always drove a transition to the all-OFF state (Figures 3I and 3J). This class of events largely explained the higher fraction of odor response in AIB neurons compared with RIM and AVA. Moreover, the AIB-only

neurons responded more rapidly to odor than AIB neurons in an all-ON state (Figure 3K). The rapid and reliable AIB-only response to odor indicates that a network property correlated with RIM and AVA activity antagonizes the AIB odor response.

To explore the functional importance of the AIB-only network state, we manipulated neuronal activity with a chemogenetic reagent, the histamine-gated chloride channel HisCl1. *C. elegans* does not use histamine as an endogenous neurotransmitter, but *C. elegans* neurons that express *Drosophila* HisCl1 are acutely hyperpolarized within a few minutes of exposure to exogenous histamine (Pokala et al., 2014). This reagent enabled the silencing of specific neurons under conditions compatible with neuronal activity imaging. HisCl1 was individually expressed in AIB, RIM, or AVA neurons, or in pairwise combinations of these neurons, to generate histamine-sensitive strains. These strains also expressed GCaMP3 in AIB, RIM, and AVA.

Acute silencing of both AVA and RIM eliminated their calcium transients, recapitulating the AIB-only state (Figure 4A). In animals in which RIM and AVA were acutely silenced with histamine, odor addition shifted AIB neurons from the ON to the OFF state with high reliability and a significantly shorter latency than controls (Figures 4B, fifth row, 4C, and 4D). A similar effect was observed when RIM and AVA were removed from the circuit using an *nmr-1::ICE* transgene that results in programmed cell death of RIM, AVA, and four other neuron classes (Zheng et al., 1999). AIB neurons in *nmr-1::ICE* animals responded rapidly and reliably to odor addition (Figure S4), resembling wild-type animals in the AIB-only state. Thus, the activity of RIM and AVA neurons delays and diminishes odor responses in AIB.

In addition to imitating endogenous network states, the cell-specific HisCl1 transgenes made it possible to generate alternative network states. Thus, with appropriate transgenes it was possible to silence only AVA or only RIM with HisCl1, although these states were not normally observed. Silencing either RIM or AVA decreased the latency and increased the efficiency of AIB odor responses, with RIM having a stronger effect (Figures 4B, third and fourth rows, 4C, and 4D). These results indicate that AIB interneurons are subject to feedback from RIM and AVA neurons and that this feedback is one source of variability in the AIB odor response.

### Interactions between AIB, RIM, and AVA Shape Odor Responses and Network States

In the All-ON state, AIB, RIM, and AVA interneurons respond to odors after a variable delay, if at all. Using HisCl silencing, we probed the effect of each of these three neurons on the odor responses of the others, alone and in combination.

Silencing either RIM or AVA increased the speed and reliability of the odor response in AVA or RIM, respectively, as well as AIB (Figures 4B–4D). Thus, RIM and AVA can each act independently to antagonize odor responses.

Silencing AIB had little effect on the onset of odor responses in RIM and AVA (Figures 4C and 4D), indicating that other inputs such as the direct AWC to AVA synapses are sufficient to drive odor responses. However, silencing AIB considerably decreased the duration of the OFF responses induced by odor in RIM and AVA (Figures 4E and 4F), indicating that AIB stabilizes odor-induced OFF states.

Silencing RIM decreased the correlation of activity between AIB and AVA neurons, suggesting that RIM has a key role in synchronizing network states (Figure 4H). Other interactions in the circuit were suggested by pairwise silencing. For example, simultaneous silencing of AIB and RIM led to a striking reduction in spontaneous AVA activity (Figure 4B, seventh row). This result suggests that both AIB and RIM provide excitatory drive to the backward command system. None of the manipulations of neuronal activity changed the bistable, switchlike behavior of the AIB, RIM, and AVA neurons.

### Chemical Synapses Mediate the Antagonistic Effects of AIB and RIM

AIB and RIM are strongly coupled to one another in the *C. elegans* wiring diagram, with bidirectional chemical synapses as well as gap junctions. RIM and AVA are connected by unidirectional chemical synapses and by gap junctions (Figure 5A). To separate the contributions of chemical synapses and gap junctions, we used cell-specific expression of tetanus toxin light chain from *Clostridium tetani* (TeTx). TeTx reduces presynaptic vesicle release by cleaving synaptobrevin/VAMP, but should spare gap junctions and neuronal excitability (Schiavo et al., 1992).

Expression of TeTx in AIB delayed odor responses in RIM and AVA and substantially decreased the length of their OFF response. Remarkably, AIB:TeTx also delayed odor responses and shortened the initial OFF duration in AIB itself, though the effect was not as strong as in RIM and AVA (Figures 5B–5D). Thus the latency and length of an AIB response to odor depends in part on AIB synaptic output, suggesting that AIB synapses oppose the antagonistic feedback that decreases AIB reliability.

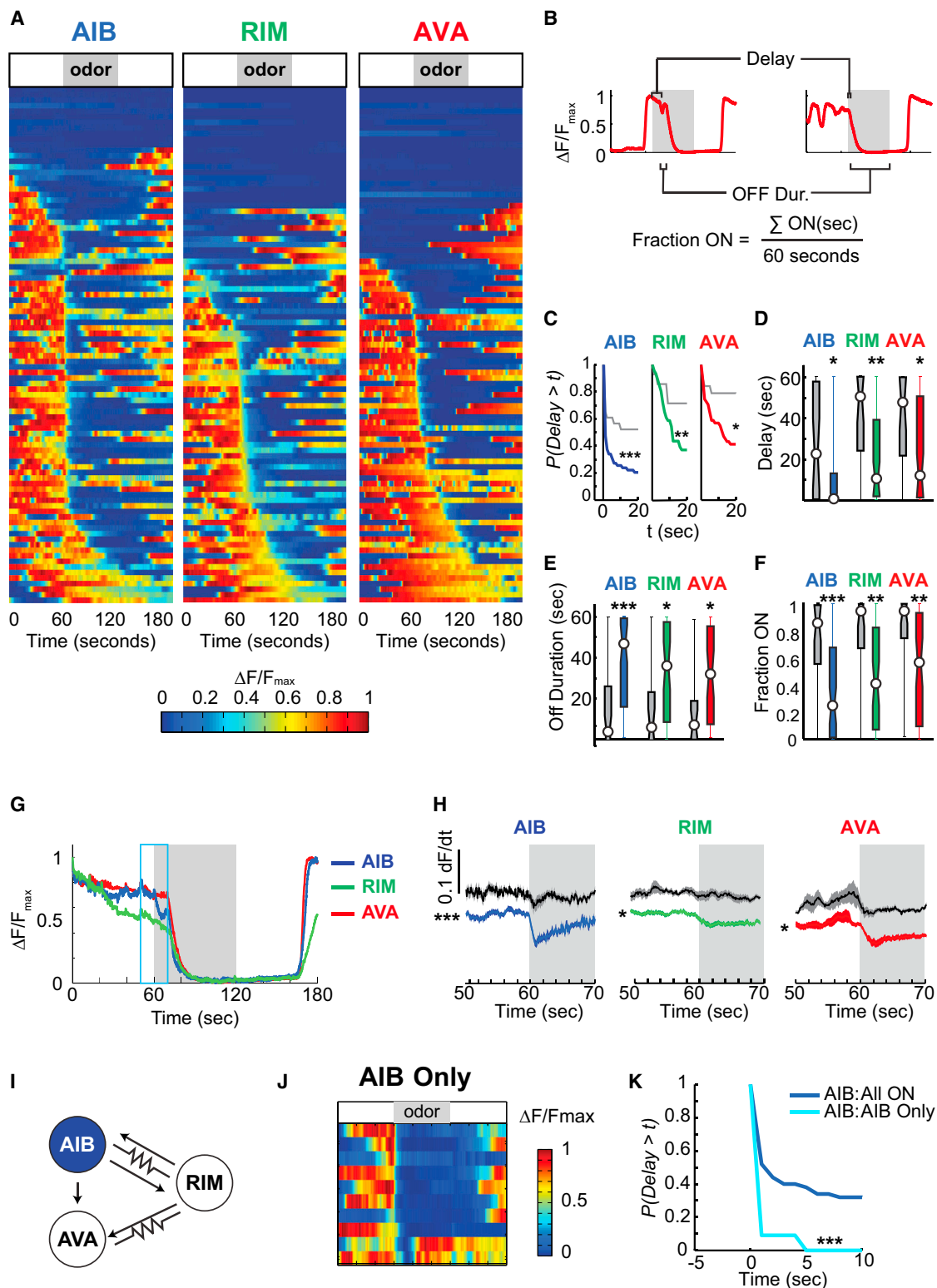
Conversely, expression of TeTx in RIM increased the reliability of AIB and AVA odor responses and decreased their latency, resembling the effects of RIM::HisCl silencing (Figures 5B–5D). RIM::TeTx also increased the reliability of the RIM response to odor. Indeed, silencing RIM chemical synapses with TeTx converted the entire network to a near-deterministic state, with over 80% of the neurons responding to odor addition with a rapid switch from the ON to OFF state that is as fast as the initial AIB response (Figures 5B–5G). These results indicate that RIM chemical synapses antagonize odor input to generate probabilistic behavior in the interneuron network.

AVA does not make chemical synapses onto AIB or RIM, but it has gap junctions with RIM. To probe the role of AVA feedback on the other neurons, we stimulated AVA with the red light-sensitive cation channel Chrimson (Klapoetke et al., 2014) during calcium imaging with GCaMP. Activation of AVA led to simultaneous calcium increase in AVA, RIM, and AIB (Figure 5H), demonstrating that feedback from AVA is sufficient to drive AIB and RIM into high-activity states.

### RIM Decreases the Reliability of Behavioral Responses to Odors

Although calcium imaging in restrained animals permitted simultaneous imaging of multiple neurons, it precluded analysis of the behaviors triggered by odor. To characterize the relationships between neuronal activity and behavior, we monitored the activity of individual interneurons in freely moving animals, using microfluidic arenas suitable for fast switching of odor stimuli





**Figure 3. AIB, RIM, and AVA Calcium Is Reduced by Odor Addition**

(A) Calcium dynamics in simultaneously recorded AIB, RIM, and AVA neurons. Traces from Figure 1D are ordered separately for each neuron according to its activity at the point of odor addition (gray bar above heatmaps) and secondarily according to the closest ON to OFF transition for that neuron. Buffer control data are in Figure S3.

(legend continued on next page)

(Larsch et al., 2013) (Figure 6A). GCaMP-expressing animals were exposed to repeated odor pulses or buffer while recording both the responses of individual interneurons and the behavior of the animal. As shown previously, AWC activity was reliably modulated by odor addition and removal (Figure 6B). On average, the activity of AIB, RIM, and AVA also fell when odor was added and rose when odor was removed, as it did in restrained animals (Figure 6B).

Alignment of neuronal activity with behavioral status of the animals showed that the activity of all three interneurons was also correlated with reversal behaviors, whether these were spontaneous (Figure 6C) or induced by odor (Figure S5). This correlation was not observed for AWC unless odor stimuli were used. The beginning of a reversal coincided with rising AIB, RIM, or AVA calcium, and the end of the reversal corresponded to falling calcium. These results suggest that the activity of AIB, RIM, and AVA is correlated in freely moving as well as restrained animals and indicate that the all-ON state corresponds to reversals. Nonetheless, there were differences in the calcium dynamics of restrained and freely moving animals; high-activity ON states were shorter in freely moving animals, as has previously been noted for AVA (Chronis et al., 2007; Ben Arous et al., 2010; Faumont et al., 2011).

Since RIM decreased the reliability of odor responses in AIB, RIM, and AVA, it should affect the corresponding behavioral responses to odor. To test this prediction, RIM synapses were inactivated with TeTx synaptic silencing, and odor pulses were delivered under conditions that led to probabilistic reversal responses (Figure 6D). As predicted by the imaging experiments, RIM:TeTx animals had a more reliable response to odor addition than wild-type animals (Figure 6D). Thus, behavioral variability in the odor response, like neuronal variability, is increased by RIM.

## DISCUSSION

Animals navigating a complex environment do not simply reproduce external stimuli through their actions; instead, they shape behavior appropriate to conditions. Chemotaxis circuits can transform smooth or noisy sensory inputs into discrete, probabilistic reorientations. More generally, animals make choices and explore environments through discrete, mutually exclusive actions, which are also the basis of experimental behavioral paradigms such as go-nogo or forced choice decisions (Frederick

et al., 2011). Here we consider one question raised by this class of behavior: how are such probabilities resolved into decisions?

Although sensory information can be limiting for performance, sensory noise in the *C. elegans* chemotaxis circuit is not the source of probabilistic behavior. AWC responds to odor rapidly and reliably in each trial, but at a behavioral level, reorientation behavior is probabilistic at odor concentrations 1,000-fold higher than the detection threshold (Larsch et al., 2013). Instead, our results indicate that interneurons integrate the sensory response with ongoing network activity to generate responses that vary in their timing and probability.

The AIB interneurons are major synaptic targets of AWC and are also direct or indirect targets of many other sensory neurons (Figure 7). Although AIB neurons respond to odors more reliably than other interneurons, they do not respond in every trial. However, AIB can respond to odors more consistently when RIM or AVA is silent. Variability in AIB response to odors thus results from interference from neurons closer to the motor response, the backward command neuron AVA and the interneuron RIM. Effectively, sensory input and RIM compete to regulate AIB, which then acts with RIM, AVA, and other neurons in a collective network state.

Both RIM and AIB chemical synapses affect the network. RIM synapses are an essential component of feedback onto AIB, and indeed the RIM connection to AIB is very strong when both synapse size and synapse number are taken into account ([www.wormwiring.org](http://www.wormwiring.org)) (Figure 7). RIM releases the neurotransmitters glutamate, acetylcholine, and tyramine, any of which could potentially act on AIB. AIB's own chemical synapses promote its odor response, potentially by antagonizing feedback from RIM and AVA. AIB uses glutamate as a transmitter, and RIM expresses both excitatory and inhibitory glutamate receptors (Hart et al., 1995; Maricq et al., 1995; Piggott et al., 2011). The detailed properties of these synapses remain to be determined. Another open question is the contribution of the gap junctions linking AIB, RIM, and AVA. Optogenetic activation of AVA can drive the network, and in the context of the wiring diagram this is likely to involve AVA gap junctions with RIM, but these connections do not have well-defined genetics or pharmacology and we did not target them directly. It should be noted that the temporal resolution of calcium imaging is limited, even when using the higher resolution provided by examining axons instead of cell bodies; therefore, this approach reports the final outcome but not the interactions between different chemical synapses and gap junctions.

(B) Example traces highlighting parameters analyzed in (C)–(F). “Delay” is the duration a neuron remained ON after odor addition. “OFF Duration” is the duration of the first OFF response after odor addition. “Fraction ON” is the fraction of time during the 1-min odor pulse that the neuron was ON.

(C–F and H) Analysis of calcium dynamics from (A) for each neuron. In all panels, responses to odors are in color, and buffer controls are in black or gray.

(C) Complementary cumulative distribution (CCD) of the ON → OFF delay for each neuron in response to a 1-min odor (red, green, blue) or buffer (gray) pulse.

(D) Length of the ON → OFF delay after odor addition.

(E) Initial OFF duration after odor addition.

(F) For neurons that were ON prior to odor exposure, the fraction of time the neuron was ON during the 1-min odor pulse. For (D)–(F), box and whisker plots show median response (circle), 25<sup>th</sup> and 75<sup>th</sup> percentile (boxes), and full distribution (lines).

(G) An example trace of AIB briefly responding to odor before a full OFF response. The blue box highlights the time frame in (H).

(H) Staggered mean time derivatives of calcium response for AIB, RIM, and AVA neurons from (A) that were ON prior to odor addition. (Buffer: AIB, n = 23; RIM, n = 24; AVA, n = 19. Odor: AIB, n = 59; RIM, n = 46; AVA, n = 51.) Shaded regions are ±SEM.

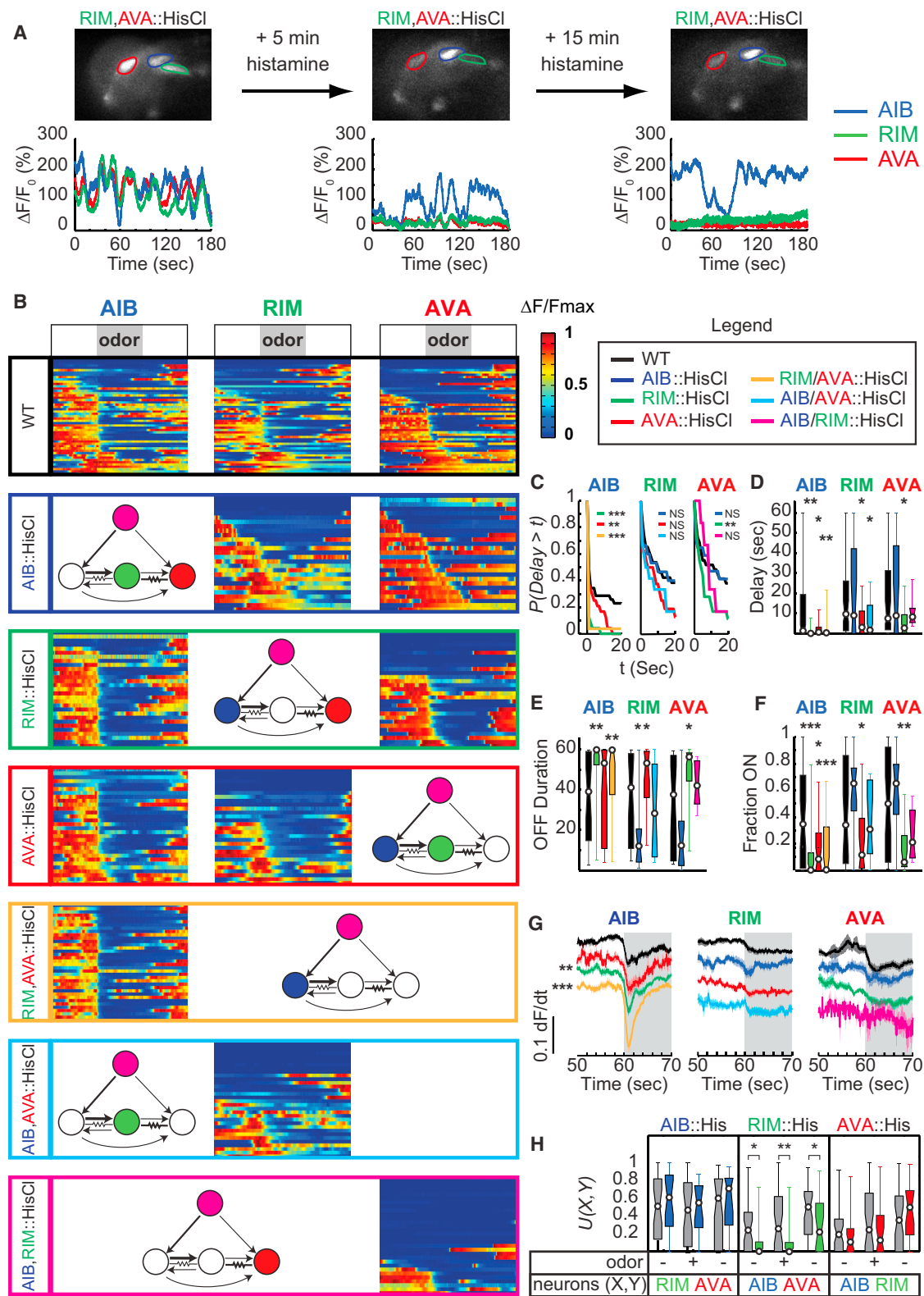
(I) Schematic diagram of the “AIB-only” state.

(J) Heatmap of AIB calcium dynamics in response to odor for AIB-only state.

(K) Complementary cumulative distribution (CCD) of the ON → OFF delay after odor addition for the AIB-only state.

p values compared with controls were calculated with a Kolmogorov-Smirnov test (C and K) or Wilcoxon rank sum test with Bonferroni correction (D–F and H).

\*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001.



**Figure 4. RIM Hyperpolarization Increases Reliability in AIB and AVA Neurons**  
(A) Representative images and traces of spontaneous activity of AIB, RIM, and AVA neurons in animals expressing HisCl1 in RIM and AVA. Left: spontaneous activity in buffer in the absence of histamine. Middle: after exposure to 10 mM histamine for 5 min. Right: after exposure to 10 mM histamine for 20 min.  
(legend continued on next page)

The *C. elegans* wiring diagram is dominated by feed-forward connections (White et al., 1986; Varshney et al., 2011), but our results demonstrate a strong feedback component from ongoing network states that evolve on a slow second to minute timescale. This feedback is evident as early as the AIB interneurons, which are generally considered sensory integrators. The faster computations at the sensory level and slower computations in collective network states allow multiple timescales of behavior to emerge. Feedback from slow network states can be envisioned as inertia in the system: once an ON or OFF state is generated, it overrides other inputs that must accumulate or wait for the network state to decay. We speculate that the collective states of AIB, RIM, and AVA represent attractor states, in which different starting points lead to a stable, self-reinforcing activity state that is either high or low (Hopfield, 1982). The factors determining the duration of the network states are unknown. Both ON and OFF states follow exponential distributions, but that leaves open a large number of mechanisms that could be stochastic, chaotic, or simply complex.

The analysis presented here provides a first view of probabilistic behavior, but there are almost certainly other neurons that participate in these decisions (Figure 7). Many other neurons are synaptically connected to AIB, RIM, and AVA and could affect transitions between high and low network states, which in turn could affect additional interacting neurons. For example, AVB, the forward command neuron, receives strong synaptic input from both AIB and RIM and synapses onto AVA; it is likely an element of an antagonistic network. Recent whole-brain imaging studies of *C. elegans* have shown that AVA, AIB, and several dozen other neurons have correlated activities (Schrödel et al., 2013; Prevedel et al., 2014). Our results agree with their conclusion that neurons in *C. elegans* have collective activity states, and the whole-brain imaging suggests that many other neurons could contribute to all-ON, all-OFF, and other possible network states.

There is still much to learn about the composition and properties of collective neuronal states, particularly in freely moving animals. Most of our results, as well as the whole-brain imaging of Schrödel et al. (2013), were obtained in the presence of a cholinergic agonist that could have altered network activity. Moreover, we and others have observed that neuronal dynamics are altered when animals are physically restrained (Ben Arous et al., 2010), perhaps because of mechanical inputs or a loss of motor and proprioceptive feedback (Kawano et al., 2011).

The importance of network state on sensory processing has long been recognized in mammalian visual cortex and other systems. For example, behavioral detection of visual stimuli by human subjects can be predicted from ongoing network activity, independent of the stimulus (Ress et al., 2000). Selective attention, task-specific information, and motor feedback can affect sensory processing in mammalian visual cortex (Silver et al., 2007; Li et al., 2004; Niell and Stryker, 2010). Motor feedback also affects the gain of visual processing in the fly brain (Maimon et al., 2010).

The integration of sensory information with network states has appealing properties from a behavioral standpoint. First, a fully determined behavior may be appropriate to certain intrinsically meaningful stimuli—highly toxic environmental conditions, food, and suitable mates—but most stimuli do not necessarily produce a predictable outcome and are not themselves rewarding. In that context, variability can prevent behavioral dead ends and increase real success. Second, different behaviors can emerge by changing the underlying state dynamics, as well as sensory properties. For example, *C. elegans* has a much higher probability of spontaneous reversals immediately after removal from food than it has an hour later, presumably accompanied by differences in AVA and RIM activity, which would lead to differences in the transmission of sensory information from AWC to AIB (Gray et al., 2005). Sensory cues that are integrated with circuit state could accordingly generate different behaviors under different feeding states. Third, adjusting the strength of different inputs in a probabilistic network increases opportunities for plasticity, and indeed AIB and RIM participate in *C. elegans* learning circuits (Ha et al., 2010). Further investigation of the generation and control of variability may provide additional insight into these more complex behavioral mechanisms.

## EXPERIMENTAL PROCEDURES

Standard culture, molecular biology, injection, and optogenetic methods were used; details and strain genotypes are in Extended Experimental Procedures. For GCaMP and HisCl expression, we used promoters for *rig-3* (expressed in AVA and some pharyngeal neurons), *tdc-1* (RIM and RIC), *inx-1* (AIB), and *str-2* (AWC<sup>ON</sup>). For ChR2 or Chrimson expression, we used the same promoters except for AVA, for which we used Cre/Lox recombination and the intersection between *nmr-1* and *rig-3* promoters (ChR2) and *glr-1* and *rig-3* promoters (Chrimson).

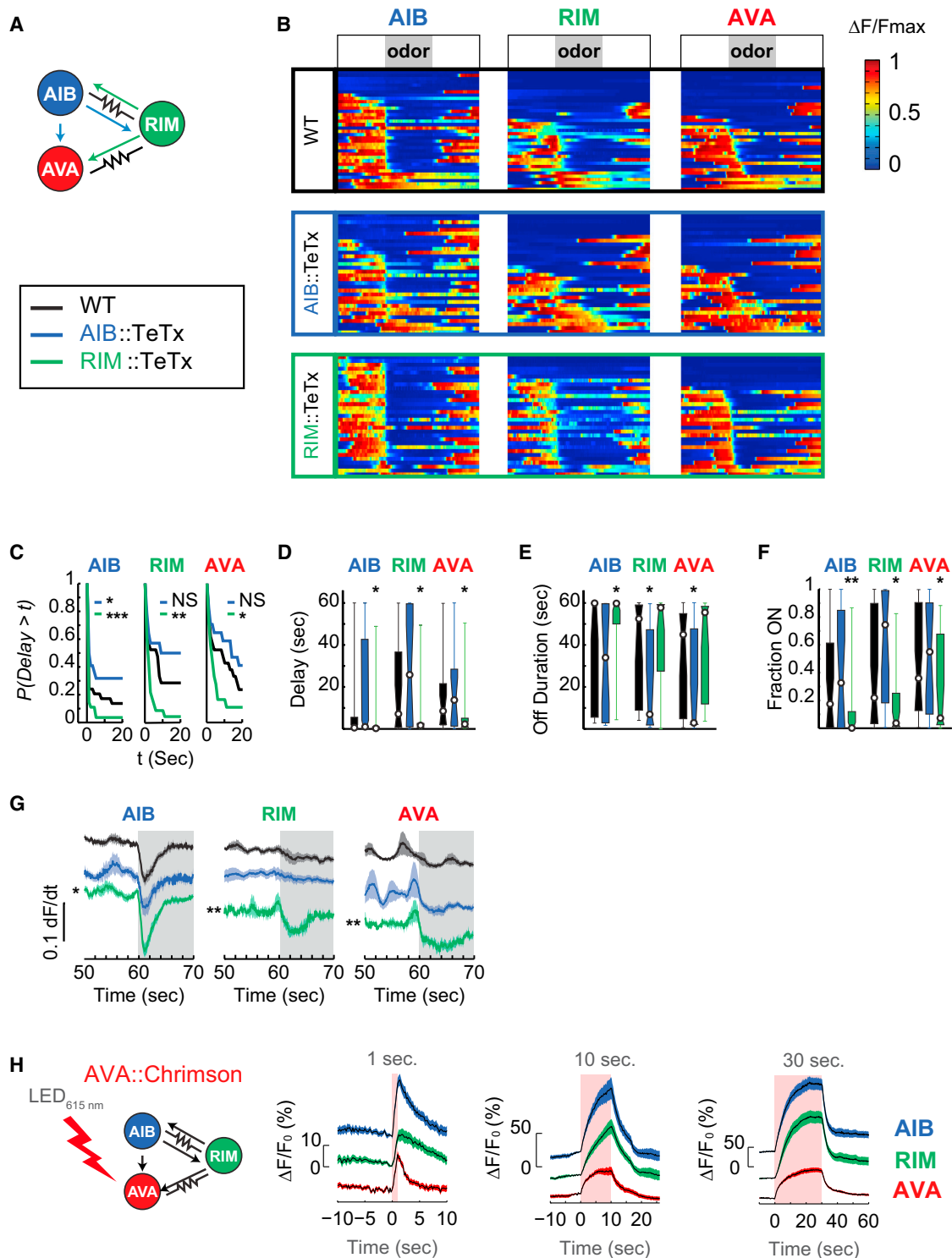
(B) Heatmaps of calcium dynamics in AIB, RIM, and AVA neurons (columns) in different HisCl-expressing strains (rows) during exposure to a 1-min pulse of 92  $\mu$ M IAA (gray bars) in the presence of histamine. Traces span 3 min and are ordered as in Figure 3A. Neuron diagrams are colored as in Figure 1A, with the histamine-silenced neurons shown in white (wild-type [WT],  $n = 51$ ; AIB::HisCl,  $n = 27$ ; RIM::HisCl,  $n = 29$ ; AVA::HisCl,  $n = 30$ ; RIM,AVA::HisCl,  $n = 36$ ; AIB,AVA::HisCl,  $n = 32$ ; AIB,RIM::HisCl,  $n = 30$ ).

(C–G) Analysis of calcium dynamics from data in (B), as in Figure 3. Color legend designates different HisCl strains. (C) Complementary cumulative distribution (CCD) of the ON  $\rightarrow$  OFF delay for each neuron in response to a 1-min odor pulse. (D) Length of the ON  $\rightarrow$  OFF delay after odor addition. (E) Initial OFF duration after odor addition. (F) For neurons that were ON prior to odor addition, the fraction of time the neuron was ON during the 1-min odor pulse. For (D)–(F) and (H), box and whisker plots show median response (circles), 25<sup>th</sup> and 75<sup>th</sup> percentile (boxes), and full distribution (lines). (G) Staggered mean time derivatives of calcium response for AIB, RIM, and AVA neurons that were ON prior to odor addition under different histamine-silenced conditions. Shaded regions are  $\pm$ SEM.

(H) Symmetric uncertainty coefficient for neuron pairs after silencing AIB, RIM, or AVA with HisCl (see Statistical Methods). A coefficient of one means both neurons are mutually dependent, while a coefficient of zero means they are completely independent. Gray indicates control values without hyperpolarization. Note that RIM and AVA are more tightly coupled to one another than to AIB. Silencing RIM significantly decreases the mutual information between AIB and AVA neurons.

p values compared with WT controls were calculated with a Kolmogorov-Smirnov test (C) or Wilcoxon rank sum test with Bonferroni correction (D–H). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .





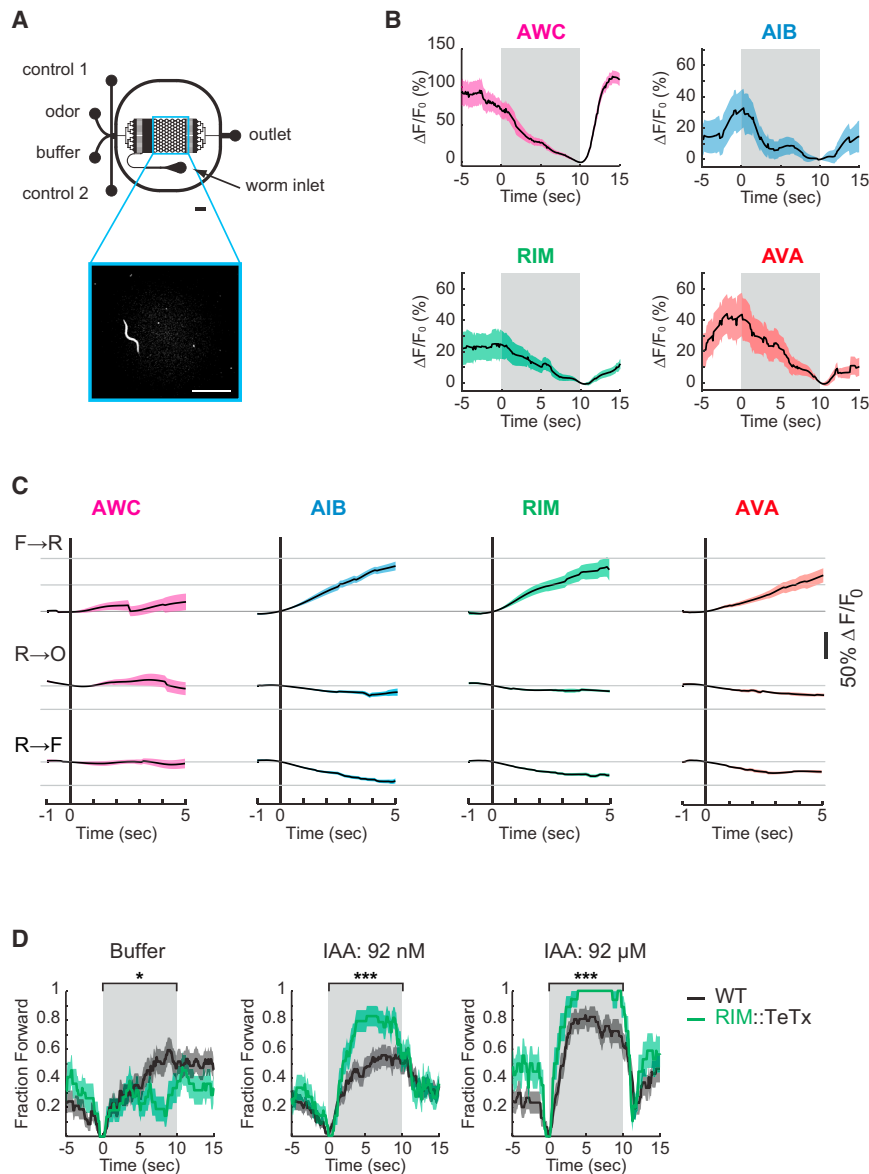
**Figure 5. RIM Chemical Synapses Drive Variability in Odor Responses**

(A) Schematic showing synapses affected by tetanus toxin expression.

(B) Heatmap of calcium dynamics in AIB, RIM, and AVA neurons (columns) in different tetanus toxin-expressing strains (rows) during exposure to a 1-min pulse of 92  $\mu$ M IAA (gray bars). Traces span 3 min and are ordered as in Figure 3A. (WT,  $n = 37$ ; AIB::TeTx,  $n = 32$ ; RIM::TeTx,  $n = 31$ ).

(C–G) Analysis of calcium dynamics from data in (B), as in Figure 3. Color legend in (A) designates TeTx-expressing strains. (C) Complementary cumulative distribution (CCD) of the ON  $\rightarrow$  OFF delay for each neuron in response to a 1-min odor pulse. (D) Length of the ON  $\rightarrow$  OFF delay after odor addition. (E) Initial OFF

(legend continued on next page)



**Figure 6. Activity of AIB, RIM, and AVA Correlates with Reversals in Freely Moving Animals**

(A) Schematic of microfluidic arena used to record calcium dynamics in freely moving animals. Animals were loaded via the worm inlet. The flow of odor or buffer into the arena was controlled by positive flow from control inlets 1 and 2, respectively. Scale bars, 1 mm.

(B) Mean neuronal calcium responses during a 10-s exposure to 92  $\mu\text{M}$  IAA (gray bar). Note different y axis for AWC and interneurons. Eight repeated 10-s exposures per animal; number of animals: AWC,  $n = 6$ ; AIB,  $n = 4$ ; RIM,  $n = 4$ ; AVA,  $n = 4$ . Shaded regions are  $\pm\text{SEM}$ .

(C) Mean neuronal calcium responses in buffer aligned to transitions between forward, reversal, and omega behaviors. Data normalized to fluorescence at point of transition ( $F_0$ ). Note common y axis for all neurons. (Forward-reversal events: AWC,  $n = 13$ ; AIB,  $n = 25$ ; RIM,  $n = 20$ ; AVA,  $n = 23$ . Reversal-omega events: AWC,  $n = 14$ ; AIB,  $n = 10$ ; RIM,  $n = 24$ ; AVA,  $n = 16$ . Reversal-forward events: AWC,  $n = 21$ ; AIB,  $n = 33$ ; RIM,  $n = 40$ ; AVA,  $n = 20$ .) Shaded regions are  $\pm\text{SEM}$ .

(D) Behavioral responses of wild-type or RIM:TeTx animals exposed to 10 s of buffer, 92 nM IAA, or 92  $\mu\text{M}$  IAA (gray bars). To match calcium imaging data, analysis included only animals showing reversals or other aversive behavior at  $t = 0$  prior to odor addition. Shaded regions are  $\pm\text{SD}$  calculated by 10,000 sampled bootstrap sampling.  $p$  values were calculated with a Kolmogorov-Smirnov test. \* $p < 0.05$ , \*\*\* $p < 0.001$ .

### Calcium Imaging in Restrained Animals

Animals were imaged in custom-built microfluidic chambers (Chronis et al., 2007) in S basal buffer (Brenner, 1974) and paralyzed with 10 mM tetramisole hydrochloride (Sigma-Aldrich) during data acquisition to reduce movement. Each animal was imaged twice, separated by a 5-min interval. For the HisCl1-silencing experiments, 10 mM histamine (Sigma-Aldrich) was included in the pre-incubation starvation conditions, the loading buffer, control buffer, and odor solutions. All animals were starved for 20 min before loading into the chip. Fluid streams were switched using a three-way valve (Lee Company). TIFF stacks were generated at 10 frames per second (fps) using a 40 $\times$  objective

(Andor iXon3 camera, Metamorph Software). In Figures 4 and 5, wild-type data were collected in parallel with genetically modified strains. Data analysis is described in Extended Experimental Procedures.

In a control dataset, the activity of AIB, RIM, and AVA was also imaged in the absence of tetramisole, although the neurons could not be imaged simultaneously due to animal movement. AIB, RIM, and AVA neurons generated spontaneous stochastic bistable activity and responded to odor probabilistically, both in the absence and presence of tetramisole (Figure S5). The immediate responses to odor addition were comparable with or without tetramisole, but the duration of the OFF response was shorter in the absence of tetramisole (Figure S5). The activity of AIB, RIM, and AVA neurons either in buffer or in odor was higher in the absence of tetramisole, i.e., the network was more often in the ON state.

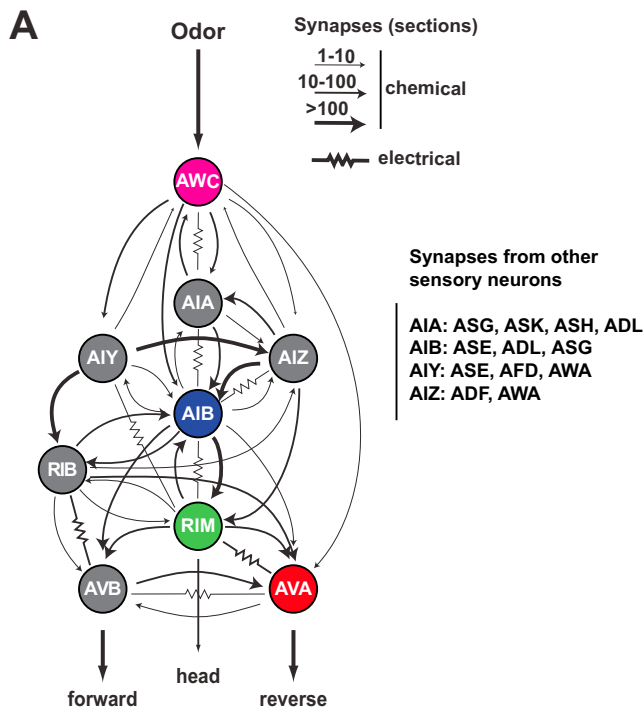
### Simultaneous Behavioral and Calcium Imaging

Methods followed those described previously (Larsch et al., 2013), using animals in a custom-fabricated 3  $\times$  3 mm polydimethylsiloxane (PDMS) imaging arena with fluid flowing through the arena by gravity flow (210-cm height

duration after odor addition. (F) For neurons that were ON prior to odor addition, the fraction of time the neuron was ON during the 1-min odor pulse. For (D)–(F), box plots show median response (circles), 25<sup>th</sup> and 75<sup>th</sup> percentile (boxes), and full distribution (lines). (G) Staggered mean time derivatives of calcium response for AIB, RIM, and AVA neurons that were ON prior to odor addition in different tetanus toxin strains.

(H) Calcium dynamics of AIB, RIM, and AVA following optogenetic stimulation of AVA by Chrimson for 1 s, 10 s, and 30 s ( $n = 28$ ). Light pulses ( $\lambda = 615$  nm) are represented by pink bar in all three panels.

Shaded regions are  $\pm\text{SEM}$ .  $p$  values compared with WT controls were calculated with a Kolmogorov-Smirnov test (C) or Wilcoxon rank sum test with Bonferroni correction (D–G). \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .



**Figure 7. Weighted Wiring Diagram for AWC, AIB, RIM, and AVA**

(A) Synaptic map of AWC, AIB, RIM, and AVA and other strongly inter-connected neurons; arrows were weighted on the basis of the total number of electron micrograph sections in which any synapse was observed (incorporates synapse size and number). Data are from [www.wormwiring.org](http://www.wormwiring.org).

differential, 15 mm/s). Each recording consisted of eight odor pulse sequences separated by 2 min of buffer exposure. Each pulse sequence lasted for 70 s, with alternate 10 s odor, 10 s buffer repeats. The first two of the eight sequences consisted of buffer-to-buffer pulses, followed by two sequences of 92 nM IAA pulses, followed by two sequences of 92  $\mu$ M IAA pulses, followed by two additional buffer-to-buffer pulses. Animals were starved in the arena for 20 min prior to data acquisition.

Metamorph software controlled the camera, light source (Lumencor SOLA-LE solid-state lamp), stimulus delivery (Automate Valvebank 8 II actuator and Lee solenoid valves), and stimulus switching (Hamilton MVP eight-way distribution valve). Stimulus switching occurred during the 2-min buffer exposure when the odor stream was bypassing the arena. TIFF images were collected at 2.5 $\times$  magnification (Hamamatsu Orca Flash 4 cMOS, Metamorph software) at 10 fps with 10-ms pulsed illumination for each 100-ms frame. Neuronal fluorescence just prior to odor removal ( $F_0$ ) was used to calculate the  $\Delta F/F_0$  for the odor-aligned fluorescence data in Figure 6B. Behaviors were binned into forward, reversal, pause and omega states, and manually corrected to ensure accuracy of the timing of transitions between states. For Figure 6C, only fluorescent data from behaviors that spanned the full observational window ( $-1$  to 5 s) were used to calculate the average.

#### Statistical Methods

Since the distribution of most data were not normally distributed based on the Shapiro-Wilk test, the significance of median differences was calculated using the Wilcoxon rank sum test with Bonferroni correction. The Kolmogorov-Smirnov test was used for comparing probability distributions. The systematic uncertainty coefficient for neuron pairs (Figure 4H) was calculated as  $(U(X, Y) = 2 [(H(X) + H(Y) - H(X, Y)) / (H(X) + H(Y))])$  (Press et al., 2002). X and Y represent the binary data for neurons X and Y,  $H(X)$  and  $H(Y)$  are the marginal entropies for X and Y, and  $H(X, Y)$  is the joint entropy for X and Y. The numerator is the mutual information for X and Y, and the denominator is the total entropy. A coefficient

of one means both neurons are mutually dependent, while a coefficient of zero means they are completely independent.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and five figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.018>.

#### AUTHOR CONTRIBUTIONS

A.G. and C.I.B. designed research; A.G. performed imaging and behavior experiments; N.P. performed optogenetic-behavior experiments; S.L. wrote tracking software for image analysis; S.W.F. developed Chrimson protocols; A.G. and C.I.B. analyzed data; and A.G. and C.I.B. wrote the paper.

#### ACKNOWLEDGMENTS

We thank L. Looger for GCaMP constructs, D. Albrecht and J. Larsch for advice with arena imaging, and S. Kato, M. Lockard, D. Ventimiglia, and other members of our laboratory for helpful discussions. This work was supported by a Marie-Josée and Henry Kravis Postdoctoral Fellowship to A.G., a Human Frontiers Science Program Fellowship to S.L., the G. Harold and Leila Y. Mathers Charitable Foundation, and the Howard Hughes Medical Institute. C.I.B. is a Howard Hughes Medical Institute investigator.

Received: October 20, 2014

Revised: January 9, 2015

Accepted: January 30, 2015

Published: March 12, 2015

#### REFERENCES

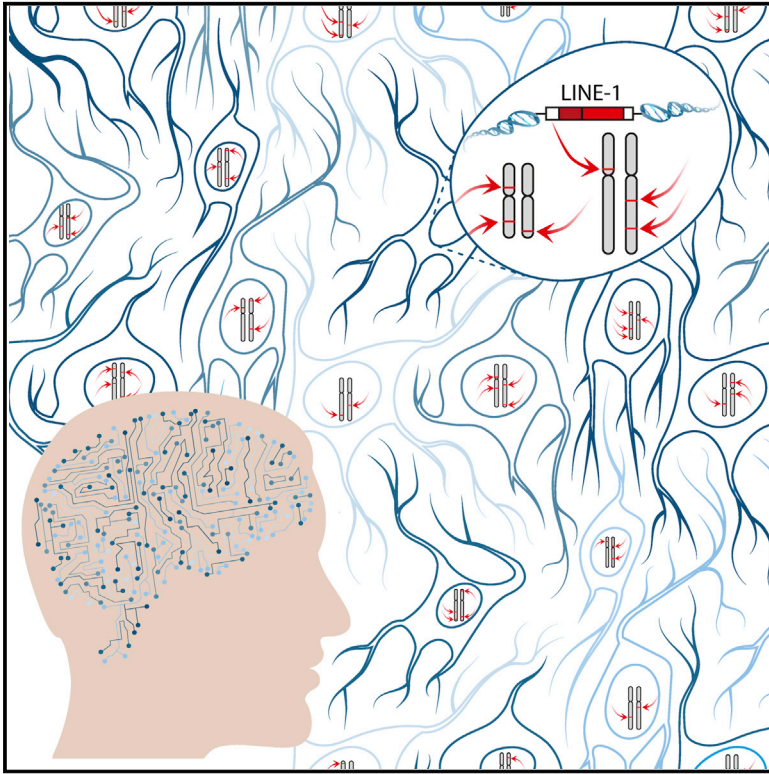
- Albrecht, D.R., and Bargmann, C.I. (2011). High-content behavioral analysis of *Caenorhabditis elegans* in precise spatiotemporal chemical environments. *Nat. Methods* 8, 599–605.
- Barlow, H.B., Levick, W.R., and Yoon, M. (1971). Responses to single quanta of light in retinal ganglion cells of the cat. *Vision Res.* S3 (Suppl 3), 87–101.
- Ben Arous, J., Tanizawa, Y., Rabinowitch, I., Chatenay, D., and Schafer, W.R. (2010). Automated imaging of neuronal activity in freely behaving *Caenorhabditis elegans*. *J. Neurosci. Methods* 187, 229–234.
- Benzi, R., Sutera, A., and Vulpiani, A. (1981). The mechanism of stochastic resonance. *J. Phys. A Math. Gen.* 14, L453–L457.
- Berg, H.C., and Brown, D.A. (1972). Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature* 239, 500–504.
- Bialek, W. (1987). Physical limits to sensation and perception. *Annu. Rev. Biophys. Biophys. Chem.* 16, 455–478.
- Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* 77, 71–94.
- Chaisanguanthum, K.S., Shen, H.H., and Sabes, P.N. (2014). Motor variability arises from a slow random walk in neural state. *J. Neurosci.* 34, 12071–12080.
- Chalasani, S.H., Chronis, N., Tsunozaki, M., Gray, J.M., Ramot, D., Goodman, M.B., and Bargmann, C.I. (2007). Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*. *Nature* 450, 63–70.
- Chalfie, M., Sulston, J.E., White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1985). The neural circuit for touch sensitivity in *Caenorhabditis elegans*. *J. Neurosci.* 5, 956–964.
- Charnov, E.L. (1976). Optimal foraging, the marginal value theorem. *Theor. Popul. Biol.* 9, 129–136.
- Chronis, N., Zimmer, M., and Bargmann, C.I. (2007). Microfluidics for in vivo imaging of neuronal and behavioral activity in *Caenorhabditis elegans*. *Nat. Methods* 4, 727–731.
- Clark, D.A., Gabel, C.V., Gabel, H., and Samuel, A.D. (2007). Temporal activity patterns in thermosensory neurons of freely moving *Caenorhabditis elegans* encode spatial thermal gradients. *J. Neurosci.* 27, 6083–6090.

- Faumont, S., Rondeau, G., Thiele, T.R., Lawton, K.J., McCormick, K.E., Sottile, M., Griesbeck, O., Heckscher, E.S., Roberts, W.M., Doe, C.Q., and Lockery, S.R. (2011). An image-free opto-mechanical system for creating virtual environments and imaging neuronal activity in freely moving *Caenorhabditis elegans*. *PLoS ONE* 6, e24666.
- Frederick, D.E., Rojas-Libano, D., Scott, M., and Kay, L.M. (2011). Rat behavior in go/no-go and two-alternative choice odor discrimination: differences and similarities. *Behav. Neurosci.* 125, 588–603.
- Goodman, M.B., Hall, D.H., Avery, L., and Lockery, S.R. (1998). Active currents regulate sensitivity and dynamic range in *C. elegans* neurons. *Neuron* 20, 763–772.
- Gray, J.M., Hill, J.J., and Bargmann, C.I. (2005). A circuit for navigation in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 102, 3184–3191.
- Guo, Z.V., Hart, A.C., and Ramanathan, S. (2009). Optical interrogation of neural circuits in *Caenorhabditis elegans*. *Nat. Methods* 6, 891–896.
- Ha, H.I., Hendricks, M., Shen, Y., Gabel, C.V., Fang-Yen, C., Qin, Y., Colón-Ramos, D., Shen, K., Samuel, A.D., and Zhang, Y. (2010). Functional organization of a neural network for aversive olfactory learning in *Caenorhabditis elegans*. *Neuron* 68, 1173–1186.
- Harsanyi, J.C. (1973). Games with randomly disturbed payoffs: a new rationale for mixed strategy equilibrium. *Int. J. Game Theory* 2, 1–23.
- Hart, A.C., Sims, S., and Kaplan, J.M. (1995). Synaptic code for sensory modalities revealed by *C. elegans* GLR-1 glutamate receptor. *Nature* 378, 82–85.
- Hendricks, M., Ha, H., Maffey, N., and Zhang, Y. (2012). Compartmentalized calcium dynamics in a *C. elegans* interneuron encode head movement. *Nature* 487, 99–103.
- Hessler, N.A., and Doupe, A.J. (1999). Social context modulates singing-related neural activity in the songbird forebrain. *Nat. Neurosci.* 2, 209–211.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* 79, 2554–2558.
- Humphries, N.E., Queiroz, N., Dyer, J.R., Pade, N.G., Musyl, M.K., Schaefer, K.M., Fuller, D.W., Brunnenschweiler, J.M., Doyle, T.K., Houghton, J.D., et al. (2010). Environmental context explains Lévy and Brownian movement patterns of marine predators. *Nature* 465, 1066–1069.
- Kawano, T., Po, M.D., Gao, S., Leung, G., Ryu, W.S., and Zhen, M. (2011). An imbalancing act: gap junctions reduce the backward motor circuit activity to bias *C. elegans* for forward locomotion. *Neuron* 72, 572–586.
- Kirkpatrick, S., Gelatt, C.D., Jr., and Vecchi, M.P. (1983). Optimization by simulated annealing. *Science* 220, 671–680.
- Klapoetke, N.C., Murata, Y., Kim, S.S., Pulver, S.R., Birdsey-Benson, A., Cho, Y.K., Morimoto, T.K., Chuong, A.S., Carpenter, E.J., Tian, Z., Wang, J., Xie, Y., Yan, Z., Zhang, Y., Chow, B.Y., Surek, B., Melkonian, M., Jayaraman, V., Constantine-Paton, M., Wong, G.K., and Boyden, E.S. (2014). Independent optical excitation of distinct neural populations. *Nat. Methods* 11, 338–346.
- Larsch, J., Ventimiglia, D., Bargmann, C.I., and Albrecht, D.R. (2013). High-throughput imaging of neuronal activity in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 110, E4266–E4273.
- Li, W., Piëch, V., and Gilbert, C.D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nat. Neurosci.* 7, 651–657.
- Lillywhite, P.G., and Laughlin, S.B. (1979). Transducer noise in a photoreceptor. *Nature* 277, 569–572.
- Longtin, A., Bulsara, A., and Moss, F. (1991). Time-interval sequences in bistable systems and the noise-induced transmission of information by sensory neurons. *Phys. Rev. Lett.* 67, 656–659.
- Maimon, G., Straw, A.D., and Dickinson, M.H. (2010). Active flight increases the gain of visual motion processing in *Drosophila*. *Nat. Neurosci.* 13, 393–399.
- Maricq, A.V., Peckol, E., Driscoll, M., and Bargmann, C.I. (1995). Mechanosensory signalling in *C. elegans* mediated by the GLR-1 glutamate receptor. *Nature* 378, 78–81.
- Mellem, J.E., Brockie, P.J., Madsen, D.M., and Maricq, A.V. (2008). Action potentials contribute to neuronal signaling in *C. elegans*. *Nat. Neurosci.* 11, 865–867.
- Mitsutake, A., Mori, Y., and Okamoto, Y. (2013). Enhanced sampling algorithms. *Methods Mol. Biol.* 924, 153–195.
- Nagel, G., Brauner, M., Liewald, J.F., Adeishvili, N., Bamberg, E., and Gottschalk, A. (2005). Light activation of channelrhodopsin-2 in excitable cells of *Caenorhabditis elegans* triggers rapid behavioral responses. *Curr. Biol.* 15, 2279–2284.
- Niell, C.M., and Stryker, M.P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65, 472–479.
- Ölveczky, B.P., Andalman, A.S., and Fee, M.S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* 3, e153.
- Pierce-Shimomura, J.T., Morse, T.M., and Lockery, S.R. (1999). The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis. *J. Neurosci.* 19, 9557–9569.
- Piggott, B.J., Liu, J., Feng, Z., Wescott, S.A., and Xu, X.Z.S. (2011). The neural circuits and synaptic mechanisms underlying motor initiation in *C. elegans*. *Cell* 147, 922–933.
- Pokala, N., Liu, Q., Gordus, A., and Bargmann, C.I. (2014). Inducible and titratable silencing of *Caenorhabditis elegans* neurons in vivo with histamine-gated chloride channels. *Proc. Natl. Acad. Sci. USA* 111, 2770–2775.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2002). Numerical recipes in C++: the art of scientific computing (Cambridge: Cambridge University Press).
- Prevedel, R., Yoon, Y.G., Hoffmann, M., Pak, N., Wetzstein, G., Kato, S., Schrödel, T., Raskar, R., Zimmer, M., Boyden, E.S., and Vaziri, A. (2014). Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat. Methods* 11, 727–730.
- Rajan, K., Abbott, L.F., and Sompolinsky, H. (2010). Stimulus-dependent suppression of chaos in recurrent neural networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 82, 011903.
- Ress, D., Backus, B.T., and Heeger, D.J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nat. Neurosci.* 3, 940–945.
- Schiavo, G., Benfenati, F., Poulain, B., Rossetto, O., Polverino de Laureto, P., DasGupta, B.R., and Montecucco, C. (1992). Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* 359, 832–835.
- Schrödel, T., Prevedel, R., Aumayr, K., Zimmer, M., and Vaziri, A. (2013). Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. *Nat. Methods* 10, 1013–1020.
- Silver, M.A., Ress, D., and Heeger, D.J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *J. Neurophysiol.* 97, 229–237.
- Thrun, S. (1992). The role of exploration in learning control. In *Handbook for Intelligent Control: Neural, Fuzzy and Adaptive Approaches*, D.A. White and D.A. Sofge, eds. (Florence, Kentucky: Van Nostrand Reinhold).
- Tian, L., Hires, S.A., Mao, T., Huber, D., Chiappe, M.E., Chalasani, S.H., Petreanu, L., Akerboom, J., McKinney, S.A., Schreier, E.R., et al. (2009). Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat. Methods* 6, 875–881.
- Tumer, E.C., and Brainard, M.S. (2007). Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* 450, 1240–1244.
- Varshney, L.R., Chen, B.L., Paniagua, E., Hall, D.H., and Chklovskii, D.B. (2011). Structural properties of the *Caenorhabditis elegans* neuronal network. *PLoS Comput. Biol.* 7, e1001066.
- White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 314, 1–340.
- Zheng, Y., Brockie, P.J., Mellem, J.E., Madsen, D.M., and Maricq, A.V. (1999). Neuronal control of locomotion in *C. elegans* is modified by a dominant mutation in the GLR-1 ionotropic glutamate receptor. *Neuron* 24, 347–361.



# Ubiquitous L1 Mosaicism in Hippocampal Neurons

## Graphical Abstract



## Authors

Kyle R. Upton, Daniel J. Gerhardt, ..., Adeline Vanderver, Geoffrey J. Faulkner

## Correspondence

faulknergj@gmail.com

## In Brief

Somatic genome mosaicism among neurons has the potential to impact brain function. L1 retrotransposons mobilize extensively in hippocampal neurons, preferentially in hippocampally expressed loci, and are depleted from mature neurons when oriented in the most deleterious configuration to host genes, suggesting functional significance.

## Highlights

- An estimated 13.7 somatic L1 insertions occur per hippocampal neuron, on average
- Target-primed reverse transcription drives somatic L1 retrotransposition
- Somatic L1 insertions sense oriented to introns are depleted in neurons and glia
- Hippocampus genes and enhancers are strikingly enriched for somatic L1 insertions



# Ubiquitous L1 Mosaicism in Hippocampal Neurons

Kyle R. Upton,<sup>1,6</sup> Daniel J. Gerhardt,<sup>1,6</sup> J. Samuel Jesuadian,<sup>1,6</sup> Sandra R. Richardson,<sup>1</sup> Francisco J. Sánchez-Luque,<sup>1</sup> Gabriela O. Bodea,<sup>1</sup> Adam D. Ewing,<sup>1</sup> Carmen Salvador-Palomeque,<sup>1</sup> Marjo S. van der Knaap,<sup>2</sup> Paul M. Brennan,<sup>3</sup> Adeline Vanderver,<sup>4</sup> and Geoffrey J. Faulkner<sup>1,5,\*</sup>

<sup>1</sup>Mater Research Institute – University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia

<sup>2</sup>Department of Child Neurology, Neuroscience Campus Amsterdam, VU University Medical Center, 1081 HV Amsterdam, The Netherlands

<sup>3</sup>Edinburgh Cancer Research Centre, Western General Hospital, Edinburgh, EH4 2XR, UK

<sup>4</sup>Center for Genetic Medicine Research, Children's National Medical Center, Washington, DC 20010, USA

<sup>5</sup>Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia

<sup>6</sup>Co-first author

\*Correspondence: [faulknergj@gmail.com](mailto:faulknergj@gmail.com)

<http://dx.doi.org/10.1016/j.cell.2015.03.026>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## SUMMARY

**Somatic LINE-1 (L1) retrotransposition during neurogenesis is a potential source of genotypic variation among neurons. As a neurogenic niche, the hippocampus supports pronounced L1 activity. However, the basal parameters and biological impact of L1-driven mosaicism remain unclear. Here, we performed single-cell retrotransposon capture sequencing (RC-seq) on individual human hippocampal neurons and glia, as well as cortical neurons. An estimated 13.7 somatic L1 insertions occurred per hippocampal neuron and carried the sequence hallmarks of target-primed reverse transcription. Notably, hippocampal neuron L1 insertions were specifically enriched in transcribed neuronal stem cell enhancers and hippocampus genes, increasing their probability of functional relevance. In addition, bias against intronic L1 insertions sense oriented relative to their host gene was observed, perhaps indicating moderate selection against this configuration *in vivo*. These experiments demonstrate pervasive L1 mosaicism at genomic loci expressed in hippocampal neurons.**

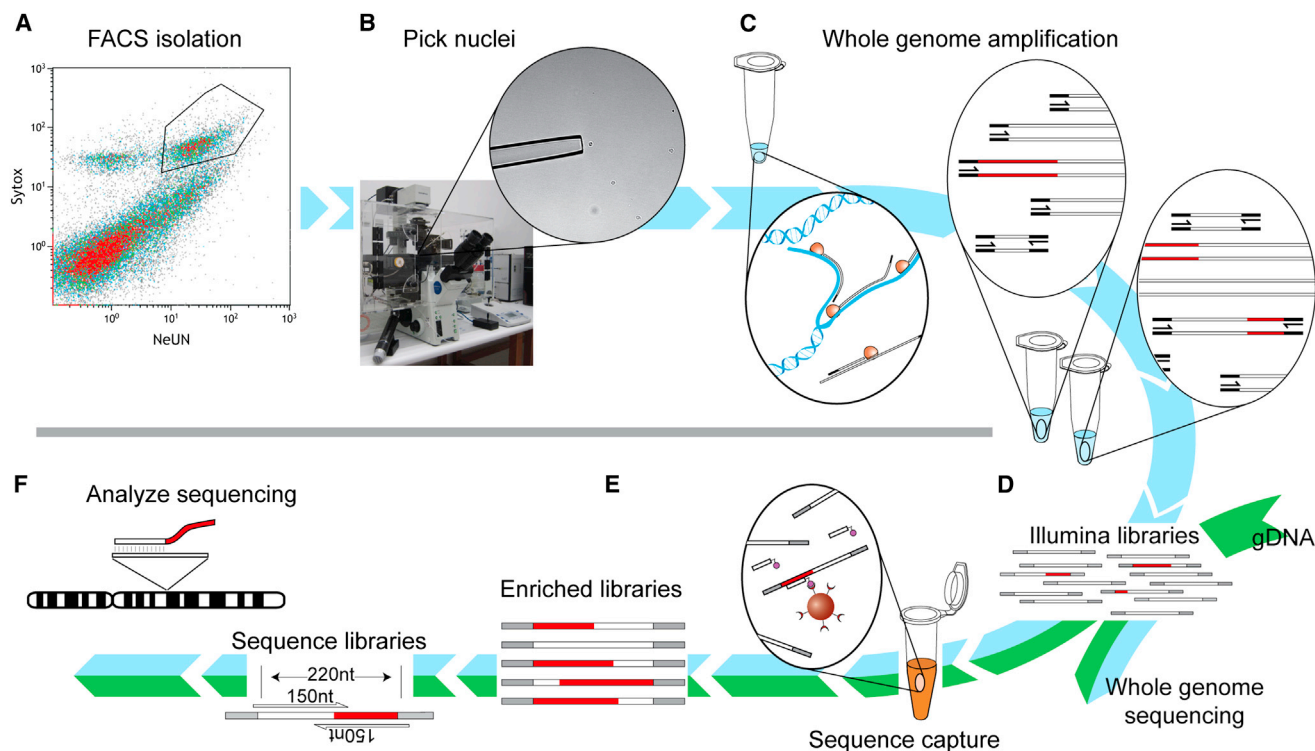
## INTRODUCTION

The extent to which the genome of one cell differs from that of any other cell from the same body is unclear. DNA replication errors, mitotic recombination, aneuploidy, and transposable element activity can cause somatic mosaicism during ontogenesis and senescence. In humans, the consequences of somatic mosaicism are most apparent in disease, including cancer and developmental syndromes (Youssoufian and Pyeritz, 2002). The impact of mosaicism among normal cells is relatively undefined beyond the notable exception of V(D)J recombination and somatic hypermutation intrinsic to lymphocyte antigen recognition (Hozumi and Tonegawa, 1976). Reports of retrotransposition (Baillie et al., 2011; Coufal et al., 2009; Evrony et al., 2012; Li et al., 2013; Muotri et al., 2005; Perrat et al., 2013) and other genomic abnormalities (Cai et al., 2014; Gole

et al., 2013; McConnell et al., 2013) in animal neurons may therefore be important given that, as for immune cells, mosaicism is a plausible route to neuron functional diversification.

Of approximately 500,000 LINE-1 (L1) copies present in the human genome, only ~100 members of the L1-Ta and pre-Ta subfamilies remain transposition-competent (Beck et al., 2010; Brouha et al., 2003). L1 mobilization primarily occurs via target primed reverse transcription (TPRT), a process catalyzed *in cis* by two proteins, ORF1p and ORF2p, translated from the bicistronic 6 kb L1 mRNA. L1 ORF2p encodes endonuclease (EN) and reverse transcriptase (RT) activities essential to L1 retrotransposition and also responsible for *trans* mobilization of *Alu* and *SVA* retrotransposons (Dewannieux et al., 2003; Hancks et al., 2011; Raiz et al., 2012). A typical TPRT-mediated L1 insertion involves a degenerate L1 EN recognition motif (5'-TT/AAAA), an L1 poly-A tail and, crucially, produces target site duplications (TSDs) (Jurka, 1997; Luan et al., 1993). Various host defense mechanisms suppress L1 activity (Beck et al., 2011), including via methylation of the CpG-rich L1 promoter. Neural progenitors and other multipotent cells can nonetheless permit L1 promoter activation (Coufal et al., 2009; Garcia-Perez et al., 2007; Wissing et al., 2012), a pattern accentuated in the hippocampus, likely due to its incorporation of the neurogenic subgranular zone (Baillie et al., 2011; Coufal et al., 2009). This coincidence of neurogenesis, L1 activity, and mosaicism has elicited speculation that L1 mobilization could impact cognitive function rooted in the hippocampus (Richardson et al., 2014).

Despite extensive evidence of somatic retrotransposition in the brain, many fundamental aspects of the phenomenon remain unclear. The rate of L1 mobilization in the neuronal lineage is, for instance, a major unresolved issue. Estimates range from <0.1 to 80 somatic L1 insertions per neuron (Coufal et al., 2009; Evrony et al., 2012). Experiments using engineered L1 reporter systems have shown that L1 mobilization is likely to occur via TPRT in neuronal precursor cells and may be altered by neurological disease (Coufal et al., 2011; Coufal et al., 2009; Muotri et al., 2005; Muotri et al., 2010). However, it is unknown whether endogenous L1 retrotransposition in hippocampal neurons adheres to these predictions. Most importantly, it is unclear whether somatic L1 insertions influence neuronal phenotype or endow carrier neuronal progenitor cells with a selective advantage or disadvantage *in vivo*. To address these questions, we



**Figure 1. Single-Cell RC-Seq Workflow**

- (A) NeuN<sup>+</sup> hippocampal nuclei were first purified by FACS (see also Figure S1).  
 (B) Nuclei were then picked using a self-contained microscope and micromanipulator.  
 (C) DNA was extracted from nuclei and subjected to linear WGA, followed by exponential PCR in two separate reactions for each nucleus, using different enzymes.  
 (D) Exponential WGA products for each nucleus were combined, used to prepare Illumina libraries, and analyzed via WGS to assess genome coverage and possible amplification biases.  
 (E) Libraries prepared in (D) were enriched via hybridization to L1-Ta LNA probes.  
 (F) Enriched libraries were sequenced with 2 × 150-mer Illumina reads and analyzed to identify novel L1 integration sites (see also Figure S2).

applied single-cell retrotransposon capture sequencing (RC-seq) to hippocampal neurons and glia, as well as cortical neurons, and found that L1 retrotransposition is a major endogenous driver of somatic mosaicism in the brain.

## RESULTS

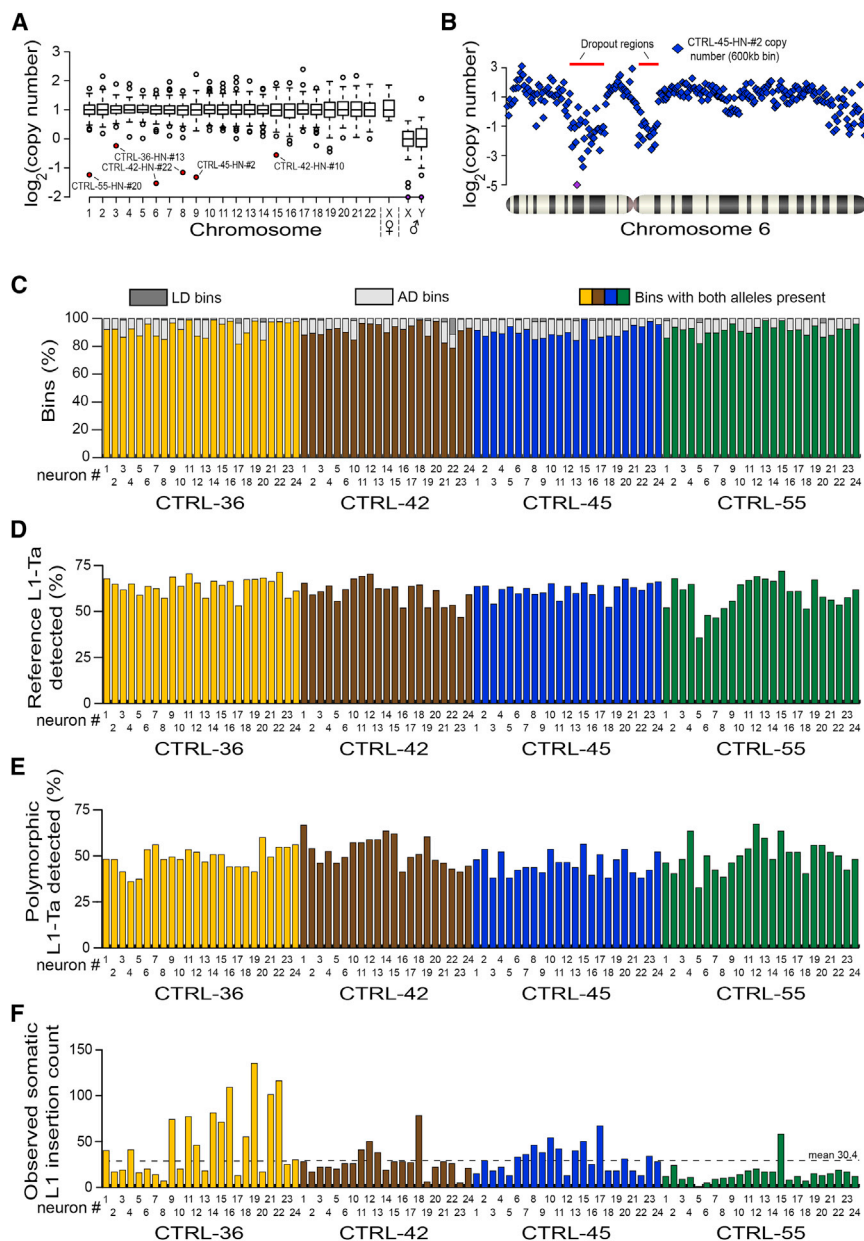
### Pervasive L1 Mobilization in Hippocampal Neurons

Several biological and technical factors hinder accurate calculation of somatic L1 mobilization frequency using bulk DNA extracted from tissue, as well as subsequent PCR validation and structural characterization of individual somatic L1 insertions (Richardson et al., 2014). We therefore developed a single-cell RC-seq protocol to detect somatic L1 insertions in individual neurons. Briefly, NeuN<sup>+</sup> hippocampal nuclei were purified by fluorescence activated cell sorting (FACS) (Figures 1A and S1), with single nuclei isolated using a self-contained microscope and micromanipulator (Figure 1B). Whole-genome amplification (WGA) was achieved through an extensively optimized version of the quasi-linear Multiple Annealing and Looping Based Amplification Cycles (MALBAC) protocol (Zong et al., 2012) and was followed by Illumina library preparation (Figures 1C and

1D). Libraries were then subjected to low-coverage (0.35×) whole-genome sequencing (WGS) as a quality control step to assess amplification bias and, in parallel, hybridized and processed by RC-seq (Figures 1E and 1F).

RC-seq utilizes sequence capture to enrich DNA for the junctions between retrotransposon termini and adjacent genomic regions, followed by paired-end sequencing, alignment, and clustering, to reveal L1 insertions absent from the reference genome. Here, we replaced previous RC-seq sequence capture pools (Baillie et al., 2011; Shukla et al., 2013) with two locked nucleic acid (LNA) probes respectively targeting the extreme 5' and 3' ends of L1-Ta. These probes capture typical L1 insertions at a 3' L1-genome junction, and full-length or heavily 5' truncated L1 insertions at a 5' L1-genome junction (Figure S2), and delivered a 15-fold improvement in L1 enrichment compared with previous RC-seq applied to brain (Baillie et al., 2011). Assembly of each overlapping read pair into a “contig” enabled computational identification of molecular chimeras and removal of PCR duplicates, and provided single-nucleotide resolution of L1 integration sites by fully spanning L1-genome junctions (Figure S2).

Prior to single-cell RC-seq, we performed deep coverage (~80×) RC-seq on bulk DNA extracted from the post-mortem



**Figure 2. Single-Cell WGS and RC-Seq Analyses of 92 Hippocampal Neurons**

(A) Chromosome copy number in each amplified genome, assessed by WGS. Box-and-whisker plots indicate median chromosomal copy number and quartiles across all neurons. Empty circles represent chromosomes with copy number  $>1.5$  IQR from the median. Sex chromosomes for CTRL-36 (female, ♀) and CTRL-42, CTRL-45, and CTRL-55 (male, ♂) are presented separately. Six autosomes, marked in red, had copy number  $\leq 1$ . Two sex chromosomes with  $\log_2$  copy number  $< -2$  are colored purple.

(B) WGS indicated 16.2 Mb and 9.4 Mb regions of localized AD (indicated by red bars) on chromosome 6 of neuron CTRL-45-HN-#2. Each blue diamond corresponds to a 600 kb “bin”. One bin with  $\log_2$  copy number  $< -5$  is colored purple.

(C) Percentages of LD (dark gray) and AD (light gray) bins in each neuron, assessed by WGS.

(D) Percentage of reference genome L1-Ta copies detected by single-cell RC-seq in each neuron.

(E) Percentage of polymorphic L1-Ta insertions found in the corresponding bulk RC-seq libraries for each individual and also detected by single-cell RC-seq.

(F) Somatic L1 insertion counts observed in each neuron by single-cell RC-seq.

Note: in (C–F) yellow, brown, blue, and green histogram columns correspond to individuals CTRL-36, CTRL-42, CTRL-45, and CTRL-55, respectively. See also [Figures S3 and S4](#) and [Tables S1 and S2](#).

hippocampus and matched liver samples of four individuals (identifiers CTRL-36, CTRL-42, CTRL-45, and CTRL-55) without evidence of neurological disease ([Table S1](#)). Bulk RC-seq on average detected 97.5% of 960 annotated reference genome L1-Ta copies ([Evrony et al., 2012](#)), indicating high assay sensitivity. As expected, we detected  $\sim 210$  polymorphic L1-Ta insertions absent from the reference genome, per individual ([Tables S1 and S2](#)). This defined the polymorphic (germline) L1-Ta insertion cohort for each individual and provided a positive control for subsequent single-cell RC-seq analyses.

Next, 92 individual neuronal nuclei were isolated from the aforementioned hippocampi, subjected to WGA and analyzed by WGS. Globally, WGS revealed that 4,226/4,232 (99.9%) chromosomes amplified ([Figure 2A](#)) with recurring WGA bias largely

limited to telomeres ([Figures S3, S4A and S4B](#)). Higher-resolution copy-number variation (CNV) analysis based on the division of the genome into adjustable-width “bins” with an average size of  $\sim 600$  kb revealed five non-telomeric deletions larger than  $\sim 5$  Mb. The largest and third largest of these occurred on chromosome 6 of CTRL-45 hippocampal neuron 2 (CTRL-45-HN-#2) and were 16.2 Mb and 9.4 Mb in length ([Figure 2B](#)). An alternative CNV analysis using  $\sim 60$  kb bins indicated the presence of numerous subregions in the 16.2 Mb example where chromosomal copy number was  $\geq 2$  ([Figure S4C](#)), depicting a region of highly variable WGA performance and, arguably, contraindicative of a genuine deletion in vivo. Genome-wide, allelic dropout (AD) and locus dropout (LD) respectively affected 8.0% and 0.7% of bins at 600 kb resolution ([Figure 2C](#), [Table S1](#)), indicating efficient amplification across  $>90\%$  of the genome. Importantly, we optimized WGA parameters to not deplete L1-Ta copies from amplified DNA, with the mean ratio of WGS reads aligned to reference L1-Ta 5′ or 3′ L1-genome junctions at 0.81 and 1.28 of expected values, respectively ([Figures S4D and S4E](#); [Table S1](#)). These results show robust WGA for individual neurons, without significant loss of reference genome L1-Ta copies.



Single-cell RC-seq applied to each of the 92 libraries analyzed by WGS detected 61.3% of reference genome L1-Ta copies (Figure 2D, Table S1) and 49.0% of polymorphic L1-Ta insertions in each neuron (Figure 2E), as defined by the earlier bulk RC-seq experiments. The latter figure provided a provisional estimate of assay sensitivity for somatic L1 insertions. A total of 2,782 putative somatic L1-Ta and pre-Ta insertions (Figure 2F, Table S2) were identified in at least one hippocampal neuron, were not detected in any bulk liver RC-seq library or more than one hippocampus by single-cell or bulk RC-seq, and were absent from existing L1 polymorphism databases (Ewing and Kazazian, 2010, 2011; Iskow et al., 2010; Shukla et al., 2013; Wang et al., 2006). Of these insertions, 1,024 (36.8%) and 34 (1.2%) were found in introns and exons, respectively. Twelve (0.4%) somatic L1 insertions were detected at both their 5' and 3' L1-genome junctions, 760 (27.3%) at only a 5' junction, and 2,010 (72.3%) at only a 3' junction. Notably, nine somatic L1 insertions detected by single-cell RC-seq were also detected and annotated as somatic in the corresponding hippocampus bulk RC-seq library, and 13 were detected by single-cell RC-seq in more than one neuron from the same hippocampus. Of somatic L1 insertions, 98.2% belonged to the L1-Ta subfamily, and 1.8% were annotated as pre-Ta. Although at 5' L1-genome junctions RC-seq captures only full-length and very heavily truncated L1s (Figure S2), we found 123 full-length L1 insertions, representing 4.4% of all events and including two instances of 5' transduction. Of those insertions detected at their 3' L1-genome junction, 151 (7.5%) carried a putative transduced 3' flanking sequence (Moran et al., 1999). This L1 3' transduction rate was lower than reported for germline L1 retrotransposition (Goodier et al., 2000), likely due to assay design not encompassing 3' transductions longer than ~100 bp, as reported elsewhere (Goodier et al., 2000; Macfarlane et al., 2013).

### PCR Validation and Structural Characterization of Somatic L1 Insertions

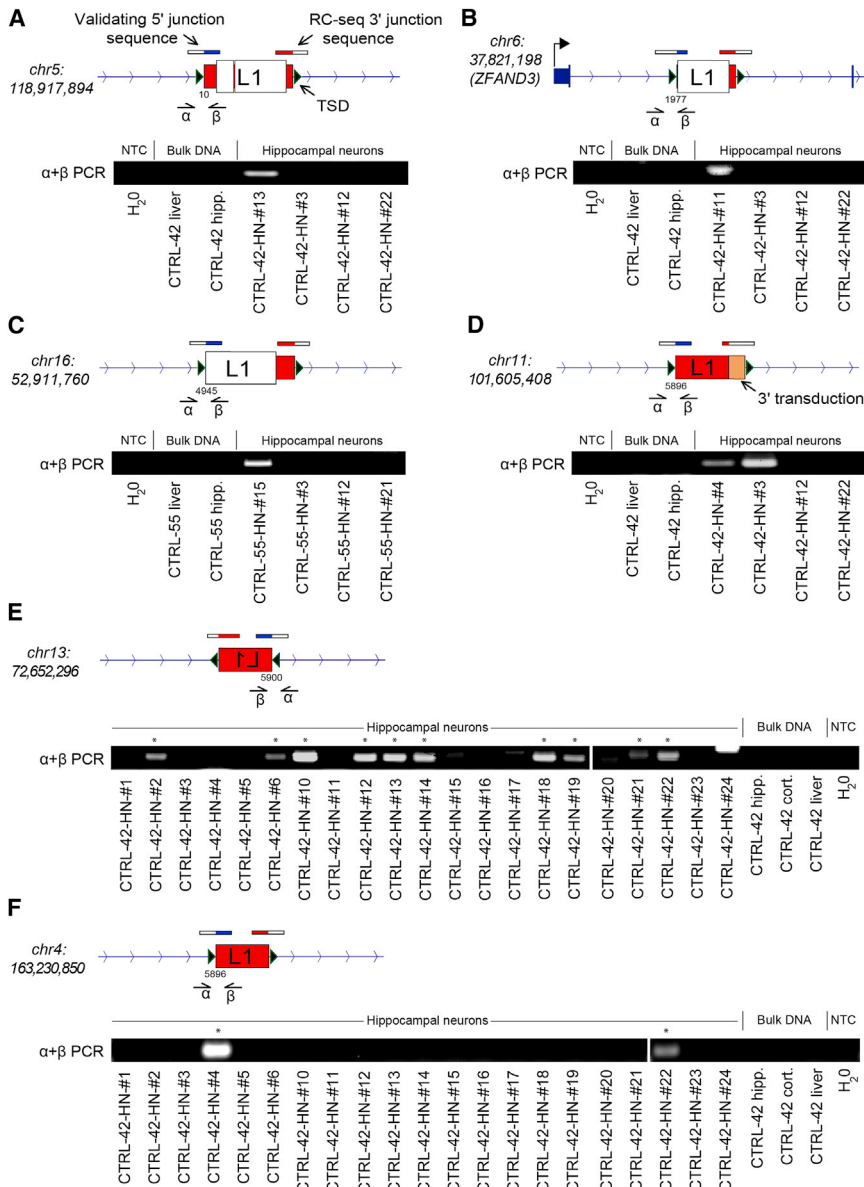
To determine the true positive rate of single-cell RC-seq, we randomly selected 20 somatic L1 insertions detected at only a 3' L1-genome junction and PCR amplified the opposing 5' L1-genome junction. This enabled detection of TPRT sequence hallmarks that distinguish WGA artifacts from most genuine L1 integration sites; specifically a TSD, an L1 EN target motif and an L1 poly-A tail (Jurka, 1997; Luan et al., 1993). Through PCR and sequencing, 5' L1-genome junctions were identified for nine insertions and, when combined with the corresponding 3' L1-genome junctions described by RC-seq, indicated TSDs and polyA-tails in all cases, and plausible L1 EN motifs for 7/9 (77.8%) examples (Tables S2 and Data S1). PCR validated insertions included full-length (Figure 3A) and variably 5' truncated (Figures 3B–F) L1s. Intronic L1 insertions were found sense oriented to two genes expressed in brain, *ZFAND3* (Figure 3B) and *USP33* (Table S2). One L1 insertion incorporated a 3' transduction and was detected by PCR in two neurons of CTRL-42 (Figure 3D). Further, PCR applied to the full panels of analyzed neurons from each individual revealed that two other L1 insertions were present in 10/21 and 2/21 neurons, respectively (Figures 3E and 3F). Three of the validated L1 insertions generated TSDs >40 bp in length.

These experiments showed that nearly half of somatic L1 insertions detected by single-cell RC-seq at a 3' L1-genome junction could be confirmed as genuine TPRT-mediated retrotransposition events. By contrast, PCR validation for 10 randomly selected exonic L1 insertions detected at a 5' L1-genome junction by single-cell RC-seq failed to find the opposing 3' L1-genome junction in all cases (Table S2). This was consistent with the L1 polyA-tail obstructing PCR amplification of somatic L1 insertion 3' ends (Baillie et al., 2011) and arguably did not resolve whether L1 insertions detected only at a 5' L1-genome junction were false positives. Finally, we selected 4 L1 insertions found at both their 5' and 3' L1-genome junctions by single-cell RC-seq; all four were confirmed by PCR and presented TPRT hallmarks, including one with a 92 bp TSD (Table S2).

Nearly 75% of somatic L1 insertions found by single-cell RC-seq were detected only at a 3' L1-genome junction (Figure S2). Given this preponderance, we sought to ascertain why the matching 5' L1-genome junction could not be identified by PCR for 11/20 selected examples of this type. PCR amplification failure was potentially due to RC-seq false positives, structurally exotic L1 insertions (Gilbert et al., 2005) or, alternatively, WGA inconsistently amplifying the 5' L1-genome junctions of insertions detected at a 3' L1-genome junction by single-cell RC-seq. To model the latter possibility, we randomly selected 12 polymorphic L1 insertions detected by bulk RC-seq and confirmed as heterozygous by genotype PCR. We performed PCR using bulk DNA to confirm each insertion was detectable at its 5' L1-genome junction and then selected 100 random examples in individual neurons where these polymorphic L1s were detected at only a 3' L1-genome junction by single-cell RC-seq (Table S2). We attempted PCR amplification of the corresponding 5' L1-genome junction for each neuron, hence recapitulating the validation process for somatic L1 insertions, and confirmed 50/100 examples. This assay indicated the maximum PCR validation rate (50.0%) for somatic L1 insertions detected at only a 3' L1-genome junction by single-cell RC-seq and, given the validation rate reported above (9/20, 45%), implied a true positive rate potentially as high as 9/10 (90.0%).

### L1 Mobilization Frequency in Diverse Neural Cell Populations

Single-cell RC-seq identified mean somatic L1 insertion counts of 48.4, 27.5, 30.5, and 14.8 per hippocampal neuron in CTRL-36, CTRL-42, CTRL-45, and CTRL-55, respectively, yielding an overall mean count of 30.4 (Figure 2F). To estimate the overall true positive mean, we incorporated the PCR validation rate (45.0%) calculated above, leading to a conservative rate calculation of 13.7 somatic L1 insertions per hippocampal neuron. If, more conservatively, only L1 insertions detected at a 3' L1-genome junction were considered, the true positive mean was 9.9. Conversely, if all L1 insertions were considered, we generously incorporated the maximum PCR validation rate calculated above (90%) and we corrected for assay sensitivity in terms of polymorphic L1 insertions detected (49.0%), the estimated true positive mean was greatly increased to 55.8. Thus, given a true positive mean of 13.7 somatic L1 insertions per neuron, and the detection of at least one event in every neuron (Figure 2F),



**Figure 3. PCR Validation of Somatic L1 Insertions**

(A–F) Validated examples from hippocampal neuron single-cell RC-seq data included: (A) a full-length L1 insertion in neuron CTRL-42-HN-#13; (B) a truncated L1 insertion in neuron CTRL-42-HN-#11; (C) a heavily truncated L1 insertion in neuron CTRL-55-HN-#15; and (D) a very heavily truncated L1 insertion yielding a 3' transduction in neuron CTRL-42-HN-#4, also validated in neuron CTRL-42-HN-#3, and traced to a donor L1-Ta on chromosome 3; (E) a very heavily truncated L1 insertion detected in CTRL-42-HN-#13 and validated in 10/21 CTRL-42 hippocampal neurons tested. Asterisks denote neurons where validation succeeded; (F) a very heavily truncated L1 insertion detected in CTRL-42-HN-#4 and also validated in CTRL-42-HN-#22. Note: in (A–F) the 3' L1-genome junction was detected by single-cell RC-seq, while the 5' L1-genome junction was identified by insertion-site PCR (using primers indicated by α and β) and sequencing. Green triangles indicate TSDs. Numbers below the 5' L1-genome junction indicate the equivalent L1-Ta consensus position. See also Table S2 and Data S1.

we concluded that L1 mosaicism was ubiquitous among the hippocampal neurons studied.

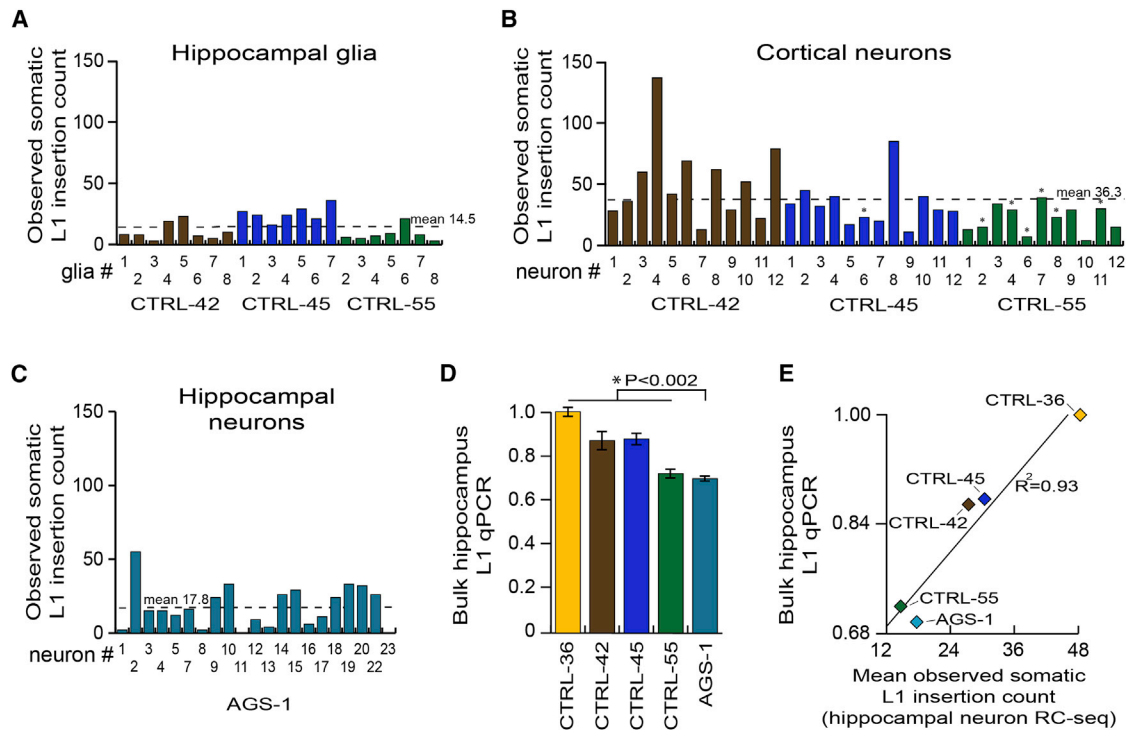
Prior *in vitro* experiments based on an engineered L1 reporter indicated that glia may support far less L1 mobilization than neurons (Coufal et al., 2009). To evaluate glial lineage endogenous L1 retrotransposition *in vivo*, we performed single-cell RC-seq upon 22 glial nuclei (NeuN<sup>−</sup>/Ki67<sup>−</sup>) isolated from CTRL-42, CTRL-45, and CTRL-55 hippocampi, and detected 316 putative somatic L1 insertions (Figures 4A and S5). This produced a mean true positive estimate of 6.5 insertions per glial cell, based on the PCR validation rate determined for hippocampal neurons (45.0%). This rate was 52.6% lower than the estimated 13.7 insertions for hippocampal neurons, a significant difference ( $p < 0.005$ , two-tailed *t* test,  $df = 112$ ). Interestingly, four insertions were found in both glial and neuronal cells by single-cell RC-

seq, with one of these instances detected at both its 5' and 3' L1-genome junctions, revealing a 12 bp TSD (Table S2). We concluded that L1 insertions can arise in proliferating neural stem cells prior to glial or neuronal commitment, while glia otherwise support less L1 mobilization than neurons.

A recent single-cell genomic analysis of 300 cortex and caudate nucleus pyramidal neurons elucidated <0.1 somatic L1 insertions per cell, and concluded that L1 was not a major driver of neuronal diversity (Evrony et al., 2012). However, the biological or technical reasons for such disparate results compared with prior data from the hippocampus were unclear. We therefore performed single-

cell RC-seq upon 35 NeuN<sup>+</sup> nuclei isolated from CTRL-42, CTRL-45 and CTRL-55 cortex tissue, including seven pyramidal neurons, and identified 1,262 putative somatic L1 insertions (Figures 4B and S5). This provided a true positive mean estimate of 16.3 insertions per cortical neuron, a figure higher than hippocampal neurons, but not significantly different. An estimated 10.7 insertions occurred per cortex pyramidal neuron, a rate substantially lower than the remaining cortical neurons but a difference that fell short of statistical significance ( $p < 0.16$ , two-tailed *t* test,  $df = 33$ ). These data elucidate L1 mosaicism in cortical neurons and exclude a biological explanation for inconsistency with the previous study.

PCR validation including TSD discovery underpins accurate calculation of L1 mobilization frequency and reflects experimental veracity independent of methodology (Richardson et al.,



**Figure 4. L1 Mobilization in Diverse Neural Cell Types**

(A) Somatic L1 insertion counts observed by single-cell RC-seq applied to hippocampal glia.

(B) As for (A) except for cortical neurons. Seven pyramidal neurons are indicated by an asterisk.

(C) As for (A) except for AGS-1 hippocampal neurons.

(D) L1 qPCR indicated lower L1 copy number in AGS-1 hippocampus versus controls ( $p < 0.002$ , two-tailed  $t$  test,  $df = 23$ ). Data represent the mean of 5 technical replicates  $\pm$  SD.

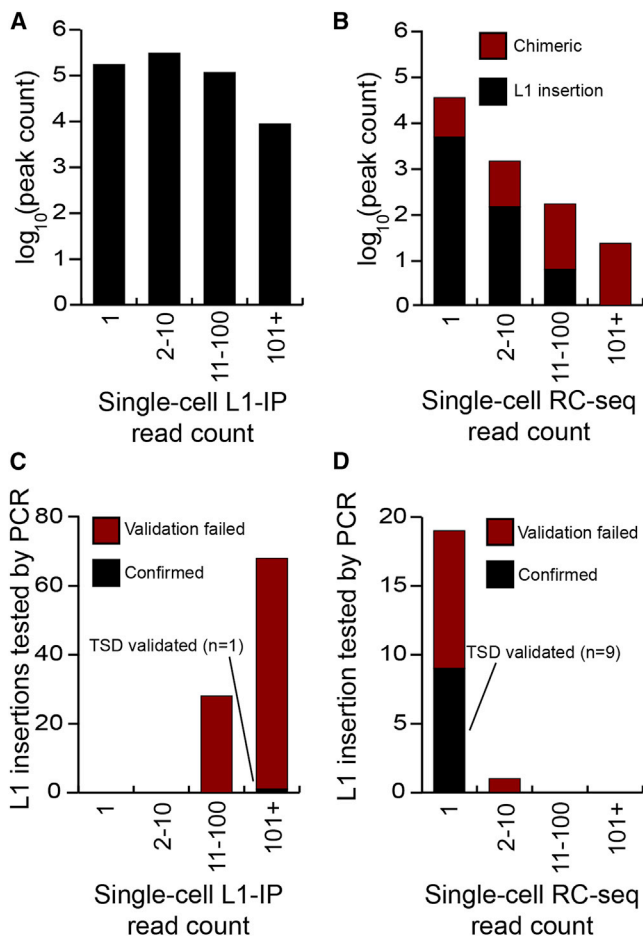
(E) Mean somatic L1 insertion counts detected by single-cell RC-seq in each hippocampus strongly correlated ( $R^2 = 0.93$ ) with L1 copy number quantified by qPCR (D).

See also [Figure S5](#) and [Table S2](#).

2014). It is therefore notable that, at this stringency, Evrony et al. reported a PCR validation rate of 1/96 and a consequential paucity of L1 activity. Two key technical considerations may explain our discrepant findings. First, RC-seq reads fully span L1-genome junctions ([Figure S2](#)), enabling bioinformatic identification of molecular chimeras before PCR validation. The earlier work by contrast followed a design ([Ewing and Kazazian, 2010](#)) that typically did not resolve L1-genome junctions, prohibiting computational removal of chimeric reads. Instead, the authors maintained that artifacts, including those generated by WGA and Illumina library preparation, should present lower read depth than genuine L1 insertions, and essentially adhered to the same principle in a very recent study applying WGS to a smaller number of neurons ([Evrony et al., 2015](#)). This assumption is crucial as, at least in single-cell RC-seq libraries, putative chimeras are disproportionately likely to amplify efficiently and accrue high read depth ([Figures 5A and 5B](#)). Second, Evrony et al. selected candidates for PCR validation effectively as a function of high read count and not at random ([Figure 5C](#)). This approach would strongly enrich for artifacts if applied to single-cell RC-seq data ([Figure 5B](#)). It follows that, without the capacity to filter artifacts a priori, the previous study resolved numerous molecular chimeras after PCR and capillary sequencing of putative L1

insertions, substantially reducing the reported validation rate. By contrast, we selected PCR validation candidates at random ([Figure 5D](#)). These factors plausibly explain why our validation rate of 9/20 (45.0%) was significantly higher than the rate of 1/96 (1.0%) reported by the earlier work ( $p < 1 \times 10^{-10}$ , chi-square test,  $df = 1$ ), as well as the disparate estimates of somatic L1 retrotransposition made by each study.

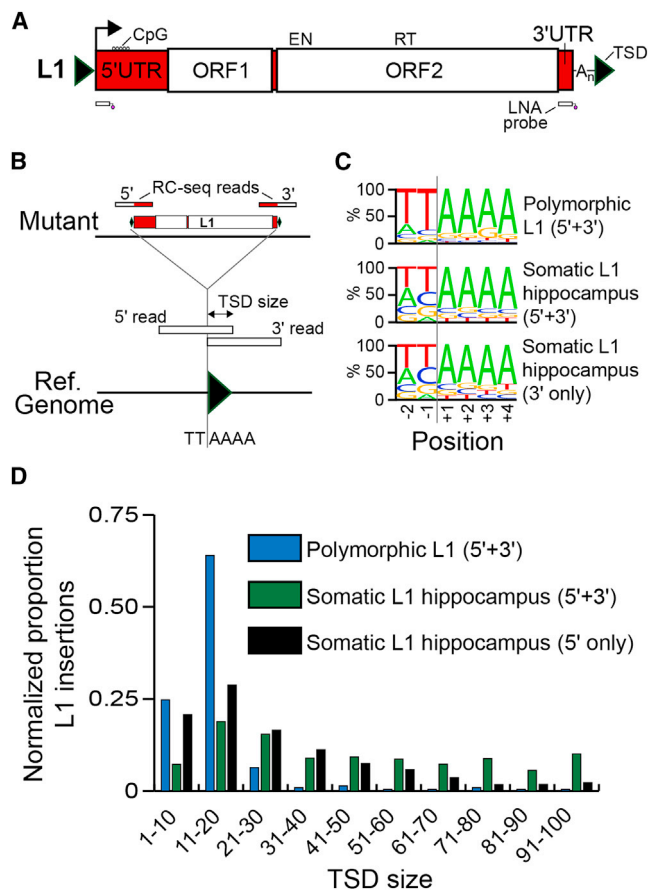
Recent qPCR based estimates of L1 CNV in human tissue, as well as in vitro L1 reporter assays, indicate L1 mobilization may be pronounced in a range of neurodevelopmental and psychiatric diseases ([Richardson et al., 2014](#)) including Aicardi-Goutières syndrome (AGS). AGS is a rare, severe neurodevelopmental condition, characterized by mutations in several genes thought to inhibit reverse transcription, including *SAMHD1* ([Zhao et al., 2013](#)). To address whether *SAMHD1* deficiency in AGS patients increases neuronal L1 mobilization, we first applied bulk RC-seq to the post-mortem hippocampus and fibroblasts of an AGS patient (identifier AGS-1) carrying two loss-of-function *SAMHD1* mutations. We then performed single-cell RC-seq upon 21 neuronal nuclei from AGS-1 hippocampus and identified 373 putative somatic L1 insertions ([Figures 4C and S5](#)), leading to a true positive mean estimate of 8.0 insertions per AGS-1 neuron. This figure was significantly ( $p < 0.03$ , two-tailed  $t$  test,  $df = 112$ ) lower



**Figure 5. Single-Cell RC-Seq Efficiently Excludes Molecular Artifacts**

(A) Distribution of read “peaks” indicating possible somatic L1 insertions detected by single-neuron L1 insertion profiling (L1-IP) (Evrony et al., 2012). (B) As for (A), except for all single-cell RC-seq data presented here. Peaks were annotated as chimeric or as likely genuine L1 insertions by sequence analysis of RC-seq reads. (C) Distribution of read peak height for L1 insertions selected for validation by Evrony et al. The L1 insertion successfully validated by TSD discovery is colored black. The remaining insertions not validated to this standard are colored red. (D) As for (C), except for L1 insertions detected by single-cell RC-seq and selected at random for validation.

than the 13.7 somatic L1 insertions found for control hippocampal neurons. A more significant difference was observed when AGS-1 neurons were compared only with the age (18 years) and gender (female) matched hippocampal neurons of CTRL-36 ( $p < 0.0001$ , two-tailed  $t$  test,  $df = 44$ ). As corollary, L1 qPCR also indicated significantly lower ( $p < 0.002$ , two-tailed  $t$  test,  $df = 23$ ) L1 copy number in AGS-1 hippocampus versus controls (Figure 4D). Finally, the results of the L1 CNV assay were strongly correlated ( $R^2 = 0.93$ ) with the mean somatic L1 insertion frequencies estimated by single-cell RC-seq (Figure 4E). We therefore concluded that L1 mobilization was unlikely to be elevated in AGS-1 hippocampus.



**Figure 6. Hallmarks of TPRT Revealed by Bulk RC-Seq**

(A) A 6 kb L1-Ta element incorporates 5' and 3' UTRs and two ORFs. ORF2p presents EN and RT domains. Methylation of a CpG island present in the 5' UTR regulates L1 promoter activity. The locations of two capture probes used by RC-seq are indicated below the L1. Note: TSDs and probes are not drawn to scale. See also Figure S2.

(B) TPRT hallmark features, including TSDs and an L1 EN recognition motif, can be identified by RC-seq, including for insertions detected at only a 5' or 3' L1-genome junction.

(C) Consensus L1 EN motifs for polymorphic and somatic L1 insertions detected at their 5' and 3' L1-genome junctions, and somatic L1 insertions found at only a 3' L1-genome junction.

(D) Observed TSD size distributions for polymorphic and somatic L1 insertions, normalized to random expectation. See also Figure S6.

### Somatic L1 Retrotransposition Occurs via TPRT

As the 13 total somatic L1 insertions detected by single-cell RC-seq and validated by PCR generally followed the TPRT model, we next assessed whether somatic L1 insertions detected by bulk RC-seq also carried TPRT signatures. RC-seq separately applied to DNA extracted from the four control hippocampus samples elucidated 318,866 putative somatic L1 insertions (Table S1). Again exploiting L1-genome junction resolution by RC-seq reads (Figures 6A and 6B and S2), we found a strong enrichment for the L1 EN motif (Figure 6C), a typical TSD size range of 5–35 nt (Figures 6D and S6) and a median L1 poly-A tail length of 33 nt for somatic L1 integration sites identified by bulk RC-seq. We also identified a substantial group of insertions



with TSDs > 40 bp in length (Figure S6). Thus, single-cell RC-seq and RC-seq applied to bulk DNA both elucidated the hallmark sequence features of TPRT-mediated retrotransposition.

### Somatic L1 Insertions Are Enriched in Neurobiology Genes

Substrate DNA chromatinization modulates L1 EN target site nicking in vitro (Cost et al., 2001). As such, dynamic changes to chromatin state during neurogenesis may impact the associated genome-wide pattern of L1 mobilization. An intersection of somatic L1 insertion sites detected by hippocampus bulk RC-seq with RefSeq gene coordinates revealed significant ( $p < 1.0 \times 10^{-150}$ , Fisher's exact test, Bonferroni correction) depletion for insertions in exons and promoters versus random sampling and significant ( $p < 3.8 \times 10^{-10}$ ) enrichment for introns versus polymorphic insertions (Table S3). Exons and introns carrying gene ontology (GO) terms relevant to neurobiology were however enriched for somatic L1 insertions (Tables S4 and S5) compared with random sampling performed by gene identifier or by genomic coordinate ( $p < 4.5 \times 10^{-5}$  and  $p < 0.03$ , respectively, Fisher's exact test, Benjamini-Hochberg correction). The latter result indicated enrichment for L1 insertions in genes expressed in the brain, despite taking into account that their length is on average >50% greater than that of other genes. By considerable margin, the most enriched GO term found (Table S5) was "regulation of synapse maturation" ( $p < 1.7 \times 10^{-60}$ , Fisher's exact test, Benjamini-Hochberg correction). Genome-wide patterns for somatic L1 insertions detected in glia and neurons by single-cell RC-seq typically corroborated those found by bulk RC-seq, including enrichment in introns and depletion from promoters and exons (Table S3) and even stronger enrichment in neurobiology genes annotated by GO term (Tables S4 and S5). Intriguingly, in AGS-1 hippocampal neurons we did not observe enrichment for L1 insertions in neurobiology genes (Table S4), whereas enrichment was observed for control hippocampal neurons, even if each individual was analyzed separately. As a control experiment, from the liver bulk RC-seq data we identified a set of 175 potential liver-specific L1 insertions (see Extended Experimental Procedures) that collectively presented a clear L1 EN consensus motif (Figure S6D) and, owing to the sensitivity of bulk RC-seq, were unlikely to represent incorrectly annotated polymorphic L1 insertions (Table S1). Notably, these liver-specific L1 insertions exhibited no enrichment for neurobiology genes (Table S4). We concluded that somatic L1 retrotransposition in neural cells preferentially occurs into the euchromatic regions of the genome contributing to neurobiology.

### Hippocampal L1 Insertions Prefer Genomic Loci Transcribed in the Hippocampus

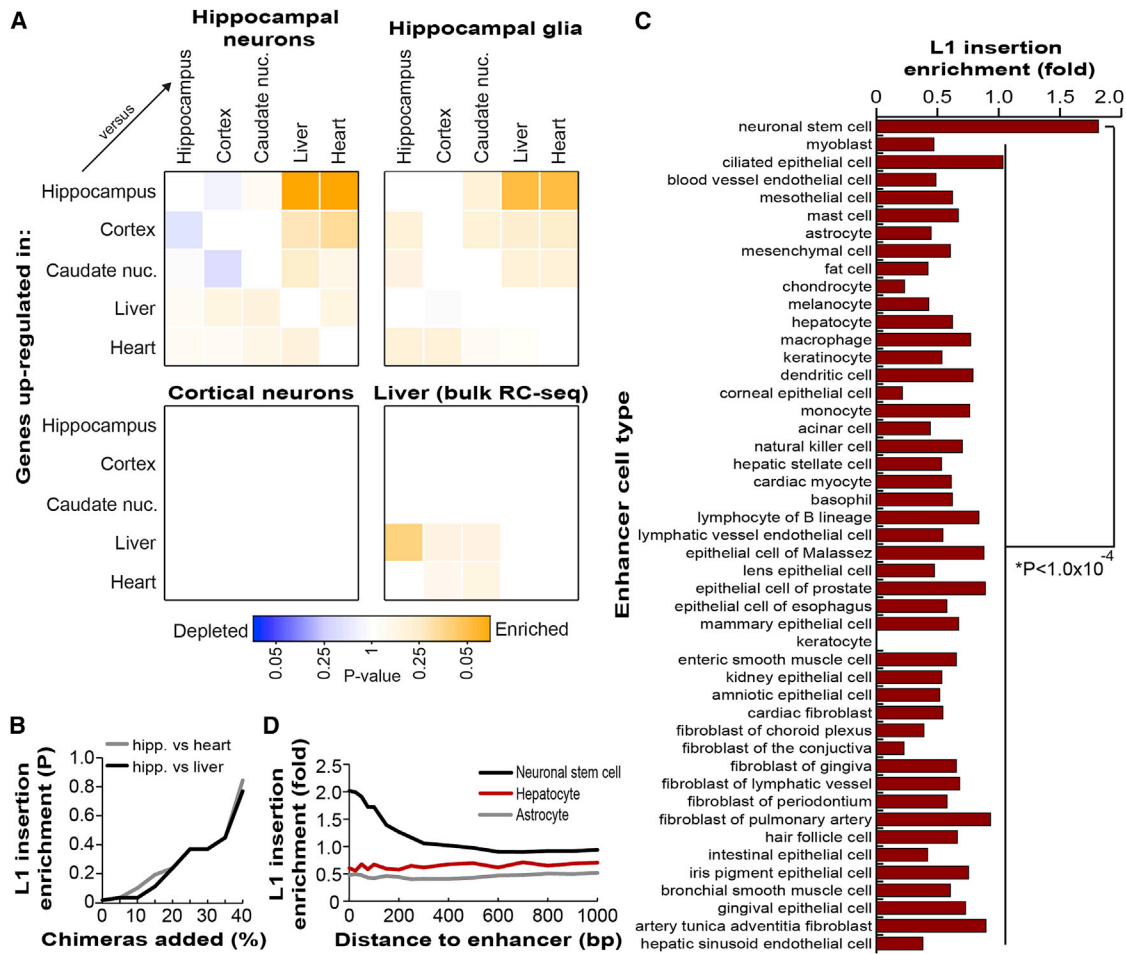
Open chromatin is a typical prerequisite for efficient transcription (Neph et al., 2012). With this in mind, we used single-molecule cap analysis of gene expression (CAGE) transcriptome profiling data from the FANTOM5 consortium (Forrest et al., 2014) to test whether genes strongly transcribed in the hippocampus were specifically enriched for somatic L1 insertions in hippocampal neurons. We first identified genes differentially upregulated in hippocampus, cortex, caudate nucleus, liver, or heart tissue surveyed by CAGE and then intersected these gene lists with the

cohort of intragenic somatic L1 insertions detected by single-cell RC-seq applied to hippocampal neurons. Only those genes upregulated in hippocampus versus heart, and hippocampus versus liver, were significantly enriched ( $p < 0.05$ , Fisher's exact test, Benjamini-Hochberg correction) for insertions (Figure 7A, Table S6). Somatic L1 insertions in hippocampal glia were also most enriched in genes upregulated in the hippocampus ( $p < 0.07$ ). No enrichment was observed for cortical neurons while, intriguingly, the liver-specific L1 insertion cohort exhibited enrichment ( $p < 0.11$ ) in genes upregulated in liver versus hippocampus (Figure 7A). Finally, we calculated the significance of enrichment for hippocampal neuron L1 insertions in genes upregulated in hippocampus while incrementally introducing putative artifacts described in Figure 5B. We found that statistical significance was no longer achieved once the dataset contained 15% or more artifacts (Figure 7B), hence demonstrating how experimental noise reduced in single-cell RC-seq analyses would otherwise obscure genome-wide enrichment. These experiments altogether reveal context-dependent, preferential L1 mobilization into strongly transcribed loci.

Noting that euchromatin is also a signature of active enhancer elements, we intersected our list of somatic L1 insertions detected by hippocampus bulk RC-seq with an extensive FANTOM5 catalog of transcribed constitutive and cell-type specific enhancers defined by histone modifications and CAGE-delineated transcriptional activity (Andersson et al., 2014). Globally, no substantial difference was observed in the rate of L1 insertions in all enhancers versus random expectation. However, of 47 cell-type specific enhancer sets, only neuronal stem cell enhancers were significantly enriched for somatic L1 insertions, compared with random expectation ( $p < 0.01$ , Fisher's exact test, Bonferroni correction) and compared with the union of the remaining 46 cell-type specific enhancer sets (Figure 7C;  $p < 1.0 \times 10^{-4}$ , Fisher's exact test). This enrichment was highest for L1 insertions within 100 nt of an enhancer, and was observed up to 500 nt from defined enhancer boundaries (Figure 7D). No enrichment was observed for astrocytes or for other cells not of the neuronal lineage, such as hepatocytes (Figure 7D). The smaller cohorts of somatic L1 insertions detected by single-cell RC-seq and liver bulk RC-seq were insufficient to perform meaningful statistical analyses of L1 insertional preference with regards to enhancers. Nonetheless, hippocampus bulk RC-seq indicated that neuronal stem cell-specific enhancers were the most highly enriched genome functional element in absolute terms (1.8-fold) for somatic L1 insertions. This reinforced the view that L1 mobilization during neurogenesis impacts regulatory and protein-coding loci specifically active in the hippocampus.

### A Potential Signature of Neurogenic L1 Selection

De novo germline L1 insertions can be highly deleterious to gene function, and commonly undergo purifying selection (Boissinot et al., 2001; Han et al., 2004). The L1 ORF2 segment of sense oriented intronic L1 insertions particularly hinders RNA polymerase processivity (Han et al., 2004; Lee et al., 2012). Hence, while sense and antisense intronic L1 insertions are assumed to occur with equal frequency in the germline, sense insertions are selected against more strongly and tend to be eliminated from



**Figure 7. Genome-Wide Somatic L1 Insertion Patterns**

(A) Somatic L1 insertions detected by single-cell RC-seq in hippocampal neurons and glia were enriched in genes differentially upregulated in hippocampus. Liver-specific L1 insertions detected by bulk RC-seq were moderately enriched in genes upregulated in liver. No enrichment was observed for cortical neurons. Color intensity is based on the absolute  $\log_2$  transformed p value determined by Fisher's exact test (Benjamini-Hochberg correction) with blue and orange colors representing depletion and enrichment, respectively. Note: in each matrix pairwise comparison, the more highly expressed tissue is on the y axis.

(B) Hippocampal somatic L1 insertions were statistically enriched in genes upregulated in hippocampus versus liver (black) or hippocampus versus heart (gray), as shown in (A). However, as previously filtered molecular chimeras (see Figure 5B) were re-introduced into this dataset, enrichment rapidly became no longer significant.

(C) Of the transcribed cell-type specific enhancers defined by FANTOM5, only those of neuronal stem cells were enriched (observed/expected) for somatic L1 insertions detected by bulk hippocampus RC-seq, compared with other enhancers ( $p < 1.0 \times 10^{-4}$ , Fisher's exact test, Bonferroni correction).

(D) Somatic L1 insertion enrichment in neuronal stem cell enhancers (black) extended 500 bp from enhancer boundaries. No enrichment was observed for astrocyte (gray) or hepatocyte (red) enhancers.

See also Tables S2, S3, S4, S5, and S6.

the population. It follows that an estimated 43.3% of recent intronic L1-Ta insertions are sense oriented, versus only 34.1% of fixed L1-Ta insertions and 39.7% of all polymorphic L1-Ta insertions (Ewing and Kazazian, 2010). By contrast, sense oriented intronic L1 insertions are not depleted in tumors (Lee et al., 2012). Among the control individuals examined here, we found that, as expected, 42/101 (41.6%) of intronic, polymorphic germline L1 insertions were sense oriented to their host gene. Surprisingly, 406/1,024 (39.6%) of intronic somatic L1 insertions detected in hippocampal neurons by single-cell RC-seq were also sense oriented, significantly less than the expected 50% ( $p < 0.0001$ , exact binomial test). This proportion was 47/136 (34.6%) and

166/503 (33.0%) for glia and cortical neurons, respectively. Adhering to the prevailing germline model of L1 evolutionary selection, we concluded that some somatic L1 insertions may arise sufficiently early in neurogenesis to impact neural progenitor cell fitness, as indicated by a depletion of sense oriented events in mature neurons and glia.

## DISCUSSION

Our experiments firmly establish that L1-driven mosaicism pervades the hippocampus and is mediated by TPRT. That we found 13.7 somatic L1 insertions per hippocampal neuron was

unexpected given a prior estimate of <0.1 insertions per cortical neuron (Evrony et al., 2012). By discovering here a myriad of L1 insertions in cortical neurons, we exclude a biological explanation for this discrepancy and instead propose that the process by which the earlier work selected insertions for validation led to a significant underestimate of L1 retrotransposition frequency. Indeed, the mobilization rate reported here much more closely resembles an earlier estimate of 80 somatic L1 insertions per brain cell, calculated via L1 qPCR (Coufal et al., 2009).

Beyond this, our data demonstrate that L1 insertions in hippocampal neurons and glia are preferentially found in protein-coding genes highly transcribed in the hippocampus. Transcribed enhancers active in neuronal stem cells are also enriched for somatic L1 insertions, indicating likely L1 perturbation of regulatory elements. L1 insertions in cortical neurons were however not significantly enriched in genes highly transcribed in the cortex. We speculate that this could be due to cortical neurogenesis primarily occurring during fetal development (Spalding et al., 2005), which presents a genome-wide transcriptional profile different to that of the adult cortex. Although L1 mobilization was not increased in AGS-1 hippocampal neurons, the pattern of L1 insertions was prospectively different to that of controls, the reasons for which are presently unclear. The most obvious caveat of this analysis is that, due to the extreme rarity of the disease, only one AGS patient hippocampus was studied. Nonetheless, this experiment serves as a proof-of-principle demonstration that single-cell RC-seq could be used in the future to assess abnormal L1 mobilization in neurological disease. Finally, we noted that somatic L1 insertions in neurons bore substantially longer TSDs on average than polymorphic L1 insertions, corroborated by structural characterization of L1 integration sites found by single-cell RC-seq. Unusually long TSDs have previously been identified using an engineered L1 reporter system in HeLa cells (Gilbert et al., 2005). As also hypothesized in that context, pervasive euchromatinization in neural progenitor cells may promote the formation of long TSDs.

The predominant developmental timing of endogenous L1 mobilization in the brain remains unclear. Although the vast majority of somatic L1 insertions detected by single-cell RC-seq were found in one cell each, a small proportion of L1s were detected in multiple cells, including examples found in both glia and neurons, indicating L1 mobilization in a common multipotent progenitor cell. Three somatic L1 insertions were validated by PCR in multiple neurons, including one example found in nearly 50% of the neurons assayed. Thus, although most L1 insertions may occur in one or a handful of neurons, a substantial number appear to arise during early neurogenesis. Indeed, the signature of potential selection against somatic L1 insertions sense oriented to host gene introns suggests that many retrotransposition events precede terminal neural cell maturation. We speculate that depletion of these events could be explained by preferential L1 integration into neurogenesis genes, thereby impacting the survival or differentiation potential of neural progenitor cells. It also cannot be excluded that somatic L1 integration primarily occurs antisense to host gene introns, though we currently lack a mechanistic explanation for this preference.

Neuronal genome mosaicism may not be restricted to somatic L1 insertions. *Alu* and SVA retrotransposons *trans* mobilized by

L1 may also contribute mosaic insertions. Other than transposable element activity, recent studies have reported localized and chromosome-wide CNV in normal neurons (Cai et al., 2014; Gole et al., 2013; McConnell et al., 2013). We find no definitive evidence of these events in our data, though it must be noted that our CNV analyses were expressly geared to discern genomic deletions caused by WGA failure or variability. However, it must be noted that we found consistent WGA inefficiency at telomeres, while others have reported that most apparent small genomic deletions occur close to telomeres (McConnell et al., 2013).

L1 mosaicism may also occur outside of the brain, for instance during early embryogenesis (Garcia-Perez et al., 2007; Kano et al., 2009) or, as we previously reported for a single L1 insertion, in the liver (Shukla et al., 2013). However, some cell types present practical and technical challenges not posed by neural cells. For example, hepatocytes are frequently multinucleated and sustain aneuploidy and polyploidy, greatly complicating single-cell genomic analysis. Thus, although the liver-specific L1 insertions detected here by bulk RC-seq consistently bore L1 EN motifs and were enriched in genes differentially upregulated in liver, we were unable to corroborate these findings with single-cell RC-seq or downstream PCR validation. Future methodological advances will therefore likely be required to elucidate L1 mosaicism in the liver, and elsewhere in the body.

The capacity to locate somatic L1 insertions in individual neural cell genomes is a major step toward determining whether mosaicism impacts neurobiological function. Limitations in assaying the transcriptome and genome of the same cell however currently prohibit functional assays of individual somatic L1 insertions. Nonetheless, given the frequency of these events, their mutagenic potential for protein-coding and regulatory regions and an apparent preference for euchromatic DNA linked to neurobiological function, it is not unreasonable to predict that L1-driven somatic mosaicism may alter the functional properties of the brain.

## EXPERIMENTAL PROCEDURES

Full protocols can be found in the [Extended Experimental Procedures](#).

### Samples

Control tissues were provided by the Edinburgh Sudden Death Brain and Tissue Bank. Tissues were obtained post-mortem from AGS-1 with ethical approval to be used as described. AGS-1 carried *SAMHD1* mutations c.646-647 delAT (p.Met216fs) and c.1223G>C (p.Arg408Pro). Patient age and gender information is provided in [Table S1](#).

### Single-Cell RC-Seq

NeuN<sup>+</sup> (neuronal) and NeuN<sup>-</sup>/Ki67<sup>-</sup> (glial) nuclei were isolated via FACS from brain tissue, individually picked under microscope and subjected to linear WGA. Products were split into three exponential PCR reactions utilizing two different kits, and then combined for library preparation and downstream PCR validation. Multiplexed Illumina libraries were pooled and sequenced (2 × 150-mer reads) to assess allelic dropout and L1-genome junction depletion, then hybridized separately to two LNA probes respectively matching the 5' and 3' ends of L1-Ta. Post-enrichment, RC-seq libraries were sequenced (2 × 150-mer reads), computationally processed, filtered to exclude artifacts, and finally used to call polymorphic and somatic L1 insertions.

### 5' L1-Genome Junction Validation and Characterization

Twenty somatic L1 insertions detected by single-cell RC-seq at a 3' L1-genome junction were selected at random for structural characterization by PCR amplification and sequencing of the corresponding 5' L1-genome junction. For each example, initial PCR template DNA consisted of WGA material from the relevant neuron. As the extent of L1 5' truncation was unknown, primers oriented antisense to L1 were designed approximately every 500 bp through the L1-Ta consensus and combined with an insertion site primer unique to each locus. 5' L1-genome junctions were identified by PCR and sequencing and then separately PCR amplified again using WGA material from the selected neuron, WGA material from other single neurons from the same individual, as well as matched bulk DNA. Amplified material was stored and handled separately to bulk DNA.

### ACCESSION NUMBERS

RC-seq and WGS data are available from the European Nucleotide Archive (ENA) using the identifier PRJEB5239.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, six tables, and one data file and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.026>.

### AUTHOR CONTRIBUTIONS

K.R.U., D.J.G., J.S.J., S.R.R., F.J.S.-L., G.O.B., A.D.E., C.S.-P., P.M.B., and G.J.F. designed and performed experiments. K.R.U. optimized single-cell WGA and analyzed RC-seq data. D.J.G. optimized RC-seq capture. J.S.J. isolated nuclei and performed WGA. P.M.B. provided control samples. M.S.v.d.K. and A.V. provided AGS patient samples. G.J.F. led the bioinformatic analyses, wrote the manuscript and directed the study. All authors commented on the manuscript.

### ACKNOWLEDGMENTS

We thank Haig Kazazian and Ryan Lister for critical reading of the manuscript. We also thank Marianna Bugiani for her work in the autopsy of AGS-1. G.J.F. acknowledges the support of an NHMRC Career Development Fellowship (GNT1045237), NHMRC Project Grants (GNT1042449, GNT1045991, GNT1067983 and GNT1068789), and the EU FP7 under grant agreement No. 259743 underpinning the MODHEP consortium. F.J.S.-L. was supported by the Alfonso Martín Escudero Foundation. P.M.B. was supported by a Wellcome Trust Clinical Fellowship (090386/Z/09/Z).

Received: September 29, 2014

Revised: December 28, 2014

Accepted: February 25, 2015

Published: April 9, 2015

### REFERENCES

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461.

Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M., et al. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534–537.

Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* **141**, 1159–1170.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* **12**, 187–215.

Boissinot, S., Entezam, A., and Furano, A.V. (2001). Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **18**, 926–935.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. USA* **100**, 5280–5285.

Cai, X., Evrony, G.D., Lehmann, H.S., Elhosary, P.C., Mehta, B.K., Poduri, A., and Walsh, C.A. (2014). Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain. *Cell Rep.* **8**, 1280–1289.

Cost, G.J., Golding, A., Schlissel, M.S., and Boeke, J.D. (2001). Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res.* **29**, 573–577.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., and Gage, F.H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127–1131.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Marchetto, M.C., Muotri, A.R., Mu, Y., Carson, C.T., Macia, A., Moran, J.V., and Gage, F.H. (2011). Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. USA* **108**, 20382–20387.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* **35**, 41–48.

Evrony, G.D., Cai, X., Lee, E., Hills, L.B., Elhosary, P.C., Lehmann, H.S., Parker, J.J., Atabay, K.D., Gilmore, E.C., Poduri, A., et al. (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* **151**, 483–496.

Evrony, G.D., Lee, E., Mehta, B.K., Benjamini, Y., Johnson, R.M., Cai, X., Yang, L., Haseley, P., Lehmann, H.S., Park, P.J., and Walsh, C.A. (2015). Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59.

Ewing, A.D., and Kazazian, H.H., Jr. (2010). High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* **20**, 1262–1270.

Ewing, A.D., and Kazazian, H.H., Jr. (2011). Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. *Genome Res.* **21**, 985–990.

Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al.; FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. *Nature* **507**, 462–470.

Garcia-Perez, J.L., Marchetto, M.C., Muotri, A.R., Coufal, N.G., Gage, F.H., O'Shea, K.S., and Moran, J.V. (2007). LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* **16**, 1569–1577.

Gilbert, N., Lutz, S., Morrish, T.A., and Moran, J.V. (2005). Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* **25**, 7780–7795.

Gole, J., Gore, A., Richards, A., Chiu, Y.J., Fung, H.L., Bushman, D., Chiang, H.I., Chun, J., Lo, Y.H., and Zhang, K. (2013). Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nat. Biotechnol.* **31**, 1126–1132.

Goodier, J.L., Ostertag, E.M., and Kazazian, H.H., Jr. (2000). Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* **9**, 653–657.

Han, J.S., Szak, S.T., and Boeke, J.D. (2004). Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268–274.

Hancks, D.C., Goodier, J.L., Mandal, P.K., Cheung, L.E., and Kazazian, H.H., Jr. (2011). Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum. Mol. Genet.* **20**, 3386–3400.

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. USA* **73**, 3628–3632.

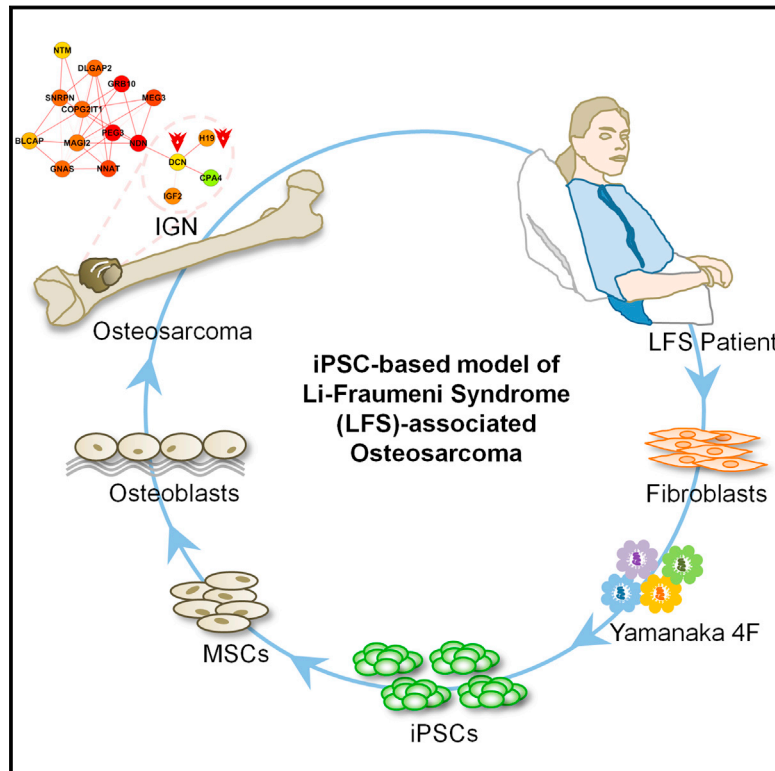
Iskrow, R.C., McCabe, M.T., Mills, R.E., Torene, S., Pittard, W.S., Neuwald, A.F., Van Meir, E.G., Vertino, P.M., and Devine, S.E. (2010). Natural



- mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261.
- Jurka, J. (1997). Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc. Natl. Acad. Sci. USA* 94, 1872–1877.
- Kano, H., Godoy, I., Courtney, C., Vetter, M.R., Gerton, G.L., Ostertag, E.M., and Kazazian, H.H., Jr. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 23, 1303–1312.
- Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., 3rd, Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K., et al.; Cancer Genome Atlas Research Network (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971.
- Li, W., Prazak, L., Chatterjee, N., Grüniger, S., Krug, L., Theodorou, D., and Dubnau, J. (2013). Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat. Neurosci.* 16, 529–531.
- Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605.
- Macfarlane, C.M., Collier, P., Rahbari, R., Beck, C.R., Wagstaff, J.F., Igoe, S., Moran, J.V., and Badge, R.M. (2013). Transduction-specific ATLAS reveals a cohort of highly active L1 retrotransposons in human populations. *Hum. Mutat.* 34, 974–985.
- McConnell, M.J., Lindberg, M.R., Brennand, K.J., Piper, J.C., Voet, T., Cowing-Zitron, C., Shumilina, S., Lasken, R.S., Vermeesch, J.R., Hall, I.M., and Gage, F.H. (2013). Mosaic copy number variation in human neurons. *Science* 342, 632–637.
- Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. *Science* 283, 1530–1534.
- Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910.
- Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446.
- Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489, 83–90.
- Perrat, P.N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., and Waddell, S. (2013). Transposition-driven genomic heterogeneity in the *Drosophila* brain. *Science* 340, 91–95.
- Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., Löwer, J., Strätling, W.H., Löwer, R., and Schumann, G.G. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res.* 40, 1666–1683.
- Richardson, S.R., Morell, S., and Faulkner, G.J. (2014). L1 retrotransposons and somatic mosaicism in the brain. *Annu. Rev. Genet.* 48, 1–27.
- Shukla, R., Upton, K.R., Muñoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111.
- Spalding, K.L., Bhardwaj, R.D., Buchholz, B.A., Druid, H., and Frisén, J. (2005). Retrospective birth dating of cells in humans. *Cell* 122, 133–143.
- Wang, J., Song, L., Grover, D., Azrak, S., Batzer, M.A., and Liang, P. (2006). dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum. Mutat.* 27, 323–329.
- Wissing, S., Muñoz-Lopez, M., Macia, A., Yang, Z., Montano, M., Collins, W., Garcia-Perez, J.L., Moran, J.V., and Greene, W.C. (2012). Reprogramming somatic cells into IPS cells activates LINE-1 retroelement mobility. *Hum. Mol. Genet.* 21, 208–218.
- Yousoufian, H., and Pyeritz, R.E. (2002). Mechanisms and consequences of somatic mosaicism in humans. *Nat. Rev. Genet.* 3, 748–758.
- Zhao, K., Du, J., Han, X., Goodier, J.L., Li, P., Zhou, X., Wei, W., Evans, S.L., Li, L., Zhang, W., et al. (2013). Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutières syndrome-related SAMHD1. *Cell Rep.* 4, 1108–1115.
- Zong, C., Lu, S., Chapman, A.R., and Xie, X.S. (2012). Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338, 1622–1626.

# Modeling Familial Cancer with Induced Pluripotent Stem Cells

## Graphical Abstract



## Authors

Dung-Fang Lee, Jie Su, ...,  
Christoph Schaniel, Ihor R. Lemischka

## Correspondence

dung-fang.lee@mssm.edu (D.-F.L.),  
ihor.lemischka@mssm.edu (I.R.L.)

## In Brief

Li-Fraumeni Syndrome patient-derived iPSCs are used to model human familial cancer, revealing a role of mutant p53 in regulating the imprinted gene network whose dysregulation results in osteoblast differentiation defects and tumorigenesis.

## Highlights

- LFS iPSC-derived OBs recapitulate OS features
- LFS iPSC-derived OBs represent OS gene signatures
- p53 mutants exert their gain-of-function effects by suppressing H19 expression
- Dysregulation of the H19 IGfN is involved in LFS-associated OS

## Accession Numbers

GSE58123



# Modeling Familial Cancer with Induced Pluripotent Stem Cells

Dung-Fang Lee,<sup>1,\*</sup> Jie Su,<sup>1,2,7</sup> Huen Suk Kim,<sup>1,2</sup> Betty Chang,<sup>1,2</sup> Dmitri Papatsenko,<sup>1</sup> Ruiying Zhao,<sup>3,8</sup> Ye Yuan,<sup>1,2</sup> Julian Gingold,<sup>1,2</sup> Weiya Xia,<sup>3</sup> Henia Darr,<sup>1</sup> Razmik Mirzayans,<sup>5</sup> Mien-Chie Hung,<sup>3,6</sup> Christoph Schaniel,<sup>1,2,4</sup> and Ihor R. Lemischka<sup>1,2,4,\*</sup>

<sup>1</sup>Department of Developmental and Regenerative Biology and The Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>2</sup>The Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>3</sup>Department of Molecular and Cellular Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA

<sup>4</sup>Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

<sup>5</sup>Department of Oncology, University of Alberta, Cross Cancer Institute, Edmonton, AB T6G 1Z2, Canada

<sup>6</sup>Center for Molecular Medicine and Graduate Institute of Cancer Biology, China Medical University, Taichung 404, Taiwan

<sup>7</sup>Present Address: Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

<sup>8</sup>Present Address: Department of Biological Sciences; Columbia University; New York, NY 10027 USA

\*Correspondence: [dung-fang.lee@mssm.edu](mailto:dung-fang.lee@mssm.edu) (D.-F.L.), [ihor.lemischka@mssm.edu](mailto:ihor.lemischka@mssm.edu) (I.R.L.)

<http://dx.doi.org/10.1016/j.cell.2015.02.045>

## SUMMARY

In vitro modeling of human disease has recently become feasible with induced pluripotent stem cell (iPSC) technology. Here, we established patient-derived iPSCs from a Li-Fraumeni syndrome (LFS) family and investigated the role of mutant p53 in the development of osteosarcoma (OS). LFS iPSC-derived osteoblasts (OBs) recapitulated OS features including defective osteoblastic differentiation as well as tumorigenic ability. Systematic analyses revealed that the expression of genes enriched in LFS-derived OBs strongly correlated with decreased time to tumor recurrence and poor patient survival. Furthermore, LFS OBs exhibited impaired upregulation of the imprinted gene H19 during osteogenesis. Restoration of H19 expression in LFS OBs facilitated osteoblastic differentiation and repressed tumorigenic potential. By integrating human imprinted gene network (IGN) into functional genomic analyses, we found that H19 mediates suppression of LFS-associated OS through the IGN component DECORIN (DCN). In summary, these findings demonstrate the feasibility of studying inherited human cancer syndromes with iPSCs.

## INTRODUCTION

Li-Fraumeni syndrome (LFS) is a genetically heterogeneous inherited cancer syndrome characterized by autosomal dominance and early onset of often multiple independent tumors within affected family members (Li and Fraumeni, 1969). In contrast to other inherited cancer syndromes predominantly characterized by site-specific cancers, LFS patients present with a variety of tumor types, including osteosarcoma (OS),

soft tissue sarcoma, breast cancer, brain tumor, leukemia, and adrenocortical carcinoma. Germline mutations in the *TP53* gene encoding the tumor suppressor p53 are responsible for LFS (Malkin et al., 1990). Mutations in p53 usually not only abolish normal p53 function but are also associated with additional oncogenic activities. Despite the prevalence of p53 mutations, the simultaneous presence of alterations in other tumor suppressors (e.g., RB1 and LKB1) and oncogenes (KRAS and HER2) makes it extremely difficult to study the specific role of p53 in cancer development. LFS provides an ideal genetic model system for investigating such a role. Although murine LFS models have been generated (Hanel et al., 2013; Lang et al., 2004; Olive et al., 2004), they do not fully recapitulate the tumor spectrum found in LFS patients. Therefore, other model systems are needed in order to further decipher mutant p53-associated pathogenesis.

Comprising almost 60% of the common histological bone sarcoma subtypes, OS is the most frequent primary non-hematological malignancy in childhood and adolescence (Tang et al., 2008). Despite advances in surgery and multi-agent chemotherapy, the survival rate has not increased in the past 40 years as much as for other malignancies. After leukemia, OS is the second leading cause of cancer mortality among children and adolescents and has been described as a cancer syndrome with a differentiation deficiency. OS exhibits osteoblast (OB)-like features and sustains undifferentiated OBs (Haydon et al., 2007). Furthermore, genetic alterations (e.g., p53 mutation and RB deletion) are strongly associated with OS development. Although the association of *TP53* mutation with OS is strongly supported by the high risk of OS in LFS patients (Porter et al., 1992), the underlying mechanism by which triggers OS development is still unclear.

H19 is a maternally imprinted gene encoding a long non-coding RNA (lncRNA). Alterations in the expression of genes in the *H19-IGF2* imprint locus are linked to both Beckwith-Wiedemann syndrome (BWS) and Russell-Silver syndrome (RSS) (Choufani et al., 2010; Eggermann, 2010). Gain of methylation of the upstream H19 imprinting center (IC1) leading to H19 inactivation

and IGF2 activation is found in 5%–10% of BWS patients and in >25% of patients with Wilms tumor, hepatoblastoma, and rhabdomyosarcoma (Choufani et al., 2010). Although the H19-IGF2 imprinting mechanism has been well documented and serves as a paradigm for the study of epigenetic regulation, the functions of H19 in biological and pathological molecular regulatory processes remain nebulous. Recently, Varrault and colleagues meta-analyzed the set of strongly correlated genes in microarray data sets to infer the “Imprinted Gene Network” (IGN), of which H19 is a member. This IGN may be part of the complex regulatory system that induces rapid but controlled growth during development (Varrault et al., 2006). H19 has been suggested to regulate embryonic growth and differentiation by controlling the expression of IGF2 and several other interconnected imprinted genes; thus, fine-tuning equilibrium of growth activation and repression (Gabory et al., 2009). These findings suggest that H19 may execute its biological functions through the IGN.

Modeling human genetic diseases has been facilitated by induced pluripotent stem cell (iPSC) methodologies (Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Yu et al., 2007). Although iPSCs are widely utilized in the study of various genetic diseases with either Mendelian or complex inheritance, their application in cancer research has been much less extensively explored. In the present study, we have modeled LFS-associated OS by using OBs derived from LFS patient-specific iPSCs and were able to recapitulate disease characteristics. The LFS iPSC-derived OBs displayed a clear OS gene expression signature whose particular transcriptional spectra strongly correlate with clinical prognosis. By integrating global transcriptional and computational analyses, we demonstrated that downregulation of H19 and its associated IGN component DECORIN (DCN) is responsible for LFS-associated OS development. Restoring H19 expression facilitates OB differentiation and inhibits tumorigenesis. Downregulation of DCN impairs H19-mediated osteogenic differentiation and tumor suppression. In summary, our results suggest that p53 mutation-mediated H19 and IGN inactivation may contribute to OS development in LFS patients and that induction of H19 expression may have important implications for the future treatment or prevention of LFS-associated OS and/or OS with somatically acquired p53 mutations.

## RESULTS

### Generation and Characterization of LFS iPSCs

To elucidate how p53 mutation results in tumor development, we generated iPSCs from patient fibroblasts obtained from a LFS family representing three LFS patients and two unaffected individuals (Figure S1A). The three patients have a heterozygous c.734G>A mutation that causes a G245D missense substitution. This site is one of the hot-spot p53 mutations in both LFS patients and somatic tumors (Varley, 2003). These patients present with a broad spectrum of tumors, including OS, neurilemmoma and astrocytoma (Figure S1A). The fibroblast samples displayed a normal karyotype under low passage (Mirzayans et al., 2010). Genome sequencing further confirmed heterozygous G245D mutations in LFS fibroblasts (Figure S1B). Using non-integrating Sendai virus (SeV)-based delivery of the four Yamanaka reprogramming factors, OCT4, SOX2, KLF4, and c-MYC (Fusaki et al.,

2009; Takahashi et al., 2007), we established a number of iPSC clones from the affected and unaffected family members. These iPSC clones all demonstrate hESC morphology and express pluripotency factors (NANOG, SOX2 and OCT4) and surface markers (TRA-1-81 and SSEA4) and alkaline phosphatase (Figure 1A). The lines also show expression of pluripotency markers at levels comparable to H9 and HES2 hESCs by quantitative (q) RT-PCR and have a more open and demethylated OCT4 promoter than the original fibroblasts (Figures 1B and 1C). We verified loss of SeV and exogenous OCT4, SOX2, KLF4, and c-MYC transgenes (Figures S1C and S1D), demonstrating that these iPSCs are zero-genetic footprint. Importantly, the iPSC lines were karyotypically normal (Figure S1E) and demonstrated the capacity to differentiate into all three germ layers in vitro (data not shown) and in teratomas (Figure 1D). All characterizations of wild-type (WT) and LFS iPSCs are summarized in Table S1. Together, these data indicate that somatic cells from LFS patients can be properly reprogrammed, maintain a pluripotent state and can be effectively differentiated.

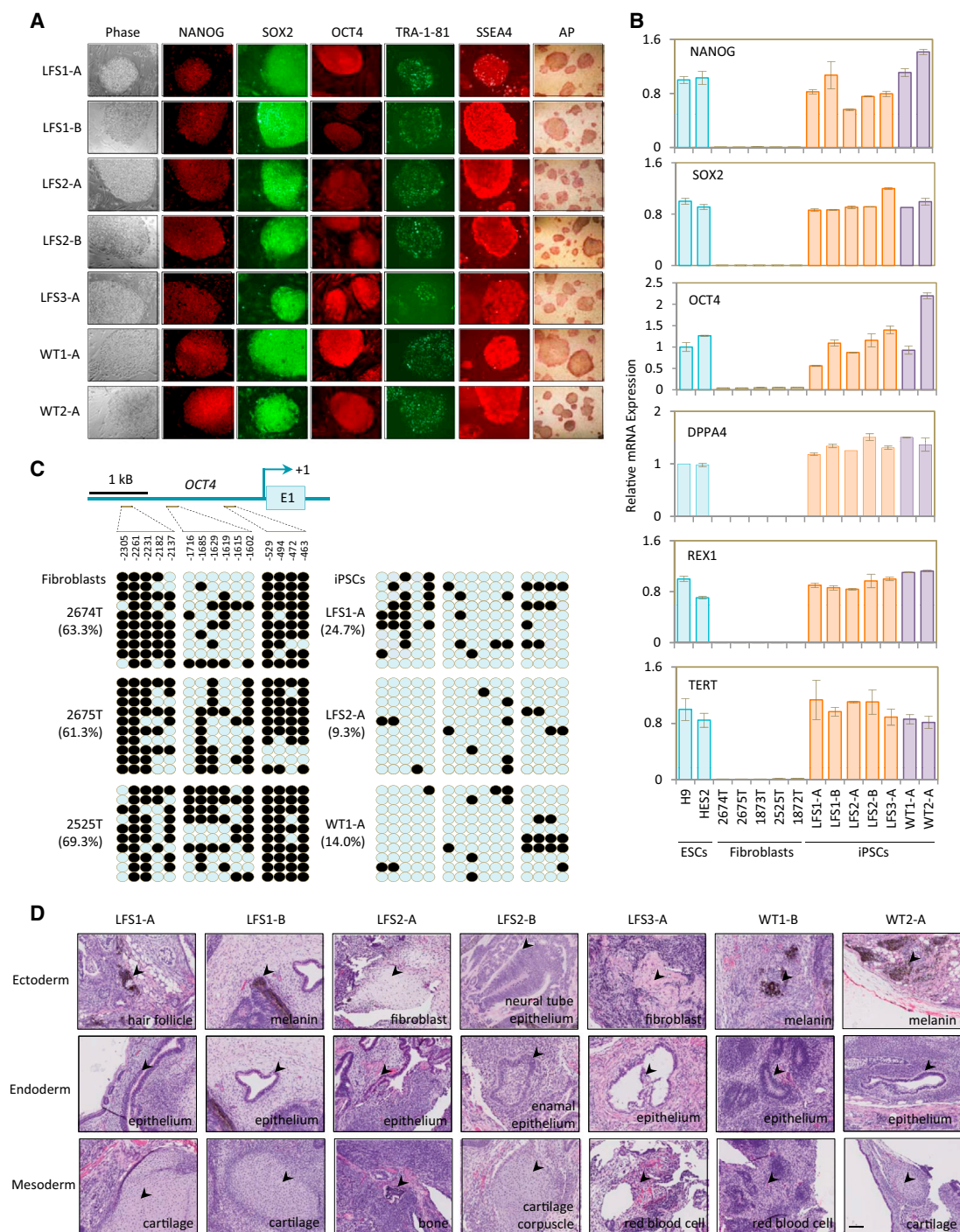
### Impairment of p53 Function in LFS iPSC-Derived Mesenchymal Stem Cells

As mentioned previously, OS, notably featuring defective OB differentiation, is one of the major cancers affecting this LFS family. Therefore, we applied our iPSC model to study how mutant p53 interferes with OB differentiation and to investigate the molecular alterations caused by p53 mutation in OS development. Human OBs can be induced from hESC-derived multipotent mesenchymal stem cells (MSCs) that can give rise to bone, cartilage, muscle, and adipose tissues. We first differentiated WT and LFS iPSCs to MSCs by treating them with FGF2 and PDGF-AB and sorting CD105<sup>+</sup>/CD24<sup>−</sup> cells (Figure S2A). These cells also expressed the MSC surface markers CD44, CD73, CD105, and CD166, and the MSC-related transcription factor SNAI1 as well as VIM (Figures 2A and S2B). The cells could be maintained for 2 months without loss of their MSC characteristics (Figure S2C) (Lian et al., 2007). In comparison with WT MSCs, LFS MSCs showed no mRNA expression differences of p53, MSC-associated transcription factors, and osteoblastic-associated factors (Figures 2B, right, S2D and S2E). Nevertheless, LFS MSCs showed lower mRNA expression levels of p53 targets p21 and MDM2 (Figure 2B, left and middle). Compared with p53(WT), p53(G245D) showed reduced binding to the p21 and MDM2 promoters by chromatin-immunoprecipitation (ChIP)-PCR analysis (Figure 2C), consistent with impaired transcriptional activity (Figure 2B). Upon MDM2 inhibitor Nutlin-3 treatment, expression of numerous p53 target genes (p21, MDM2, SFN, NOXA, FAS, TNFRSF10B, and GADD45A) were upregulated in WT MSCs, but this effect was blunted in LFS MSCs (Figure 2D). All characteristics of WT and LFS MSCs are summarized in Table S1. These studies demonstrate that LFS MSCs not only maintain MSC characteristics identical to WT MSCs but also retain the defective p53 function of the parental fibroblasts (Barley et al., 1998).

### Recapitulating OS Characteristics in LFS MSC-Derived OBs

Since it was previously suggested that impairment of p53 function leads to OS (Walkley et al., 2008) and clinical OS samples





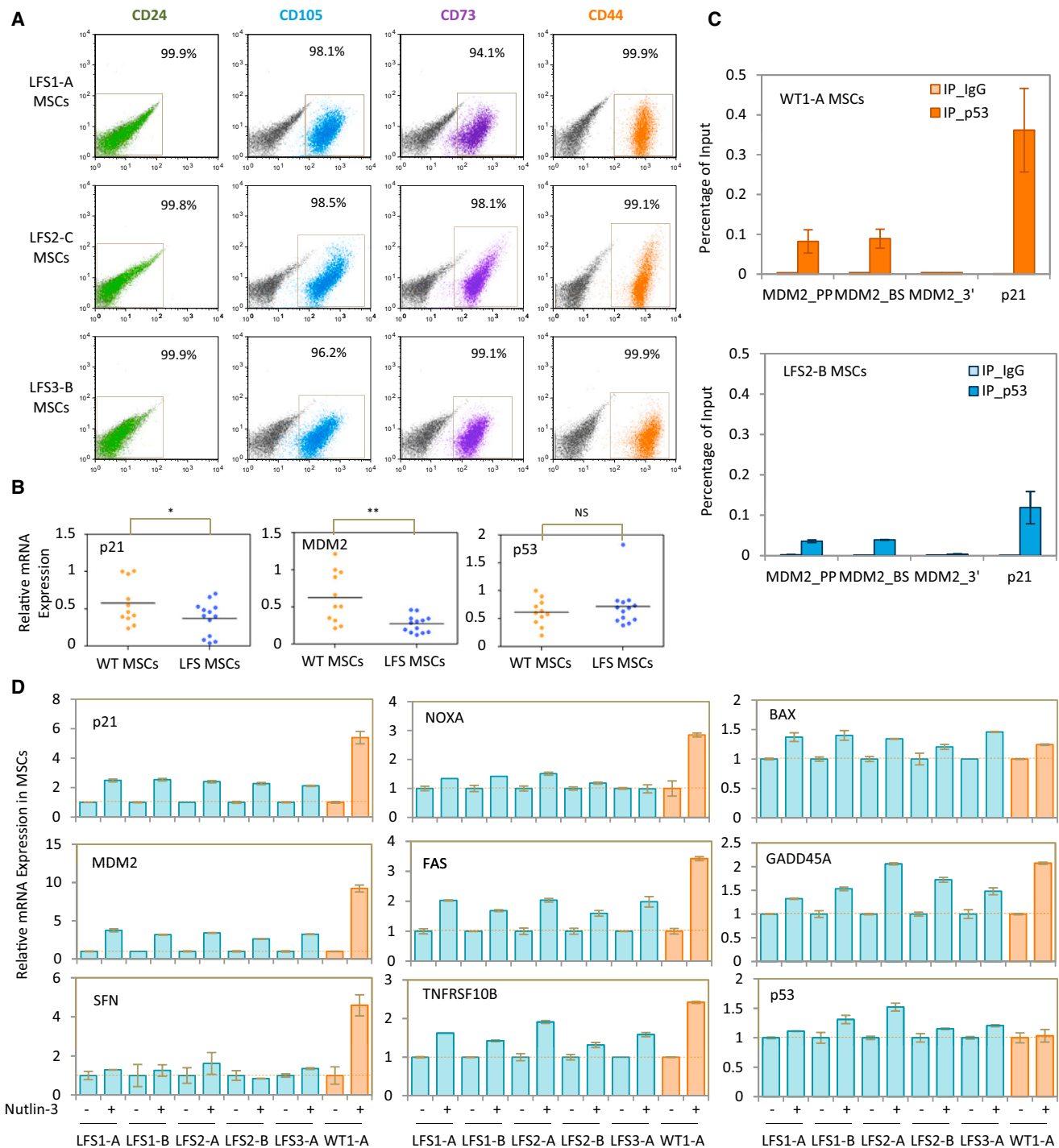
### Figure 1. LFS iPSCs Are Pluripotent

(A) SeV-4F (OCT4, SOX2, KLF4 and c-MYC) reprogrammed LFS and wild-type (WT) iPSCs derived from the same family express hESC pluripotency factors, hESC surface markers, and AP activity.

(B) qRT-PCR assay for expression of endogenous human *NANOG*, *SOX2*, *OCT4*, *DPPA4*, *REX1*, and *TERT* in iPSCs and parental fibroblasts. PCR reactions are normalized to *GAPDH* and plotted relative to expression levels in human H9 ESCs. Error bars indicate  $\pm$  SEM of triplicates.

(C) Bisulfite sequencing analysis of the *OCT4* promoter showing CpG hypomethylation in WT and LFS iPSCs relative to the parental fibroblasts. The cell line and percentage of CpG methylation are indicated to the left of each cluster. Closed circle, methylated CpG; open circle, unmethylated CpG.

(D) In vivo teratoma formation assay demonstrates LFS iPSC capacity to differentiate into the three germ layers. H&E-stained teratomas containing embryonic tissues all three germ layers, including enamel epithelium (endoderm); neural tube epithelium (ectoderm); cartilage corpuscle (mesoderm). Scale bar, 100  $\mu$ m. See also Figure S1 and Table S1.



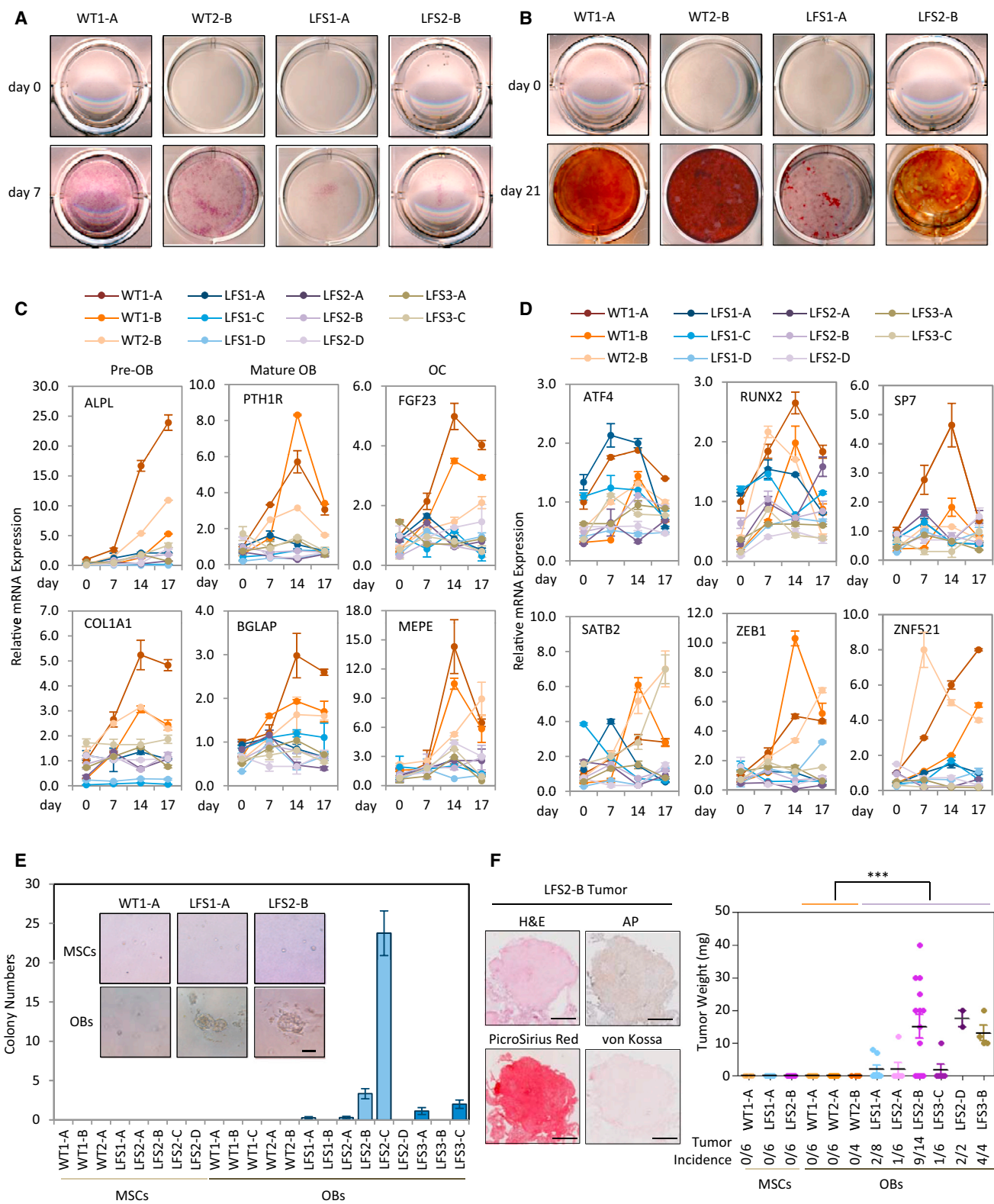
**Figure 2. Defective p53 Activity in LFS MSCs**

(A) Surface antigen profiling of LFS MSCs by flow cytometry demonstrating the CD24<sup>+</sup>, CD105<sup>+</sup>, CD73<sup>+</sup> and CD44<sup>+</sup> fractions in differentiated MSCs.

(B) qRT-PCR analysis for expression of p53 and its downstream target genes, p21, and MDM2, in the WT MSC group (11 lines) and the LFS MSC group (13 lines). The p53 mRNA levels do not show a significant difference between WT and LFS MSCs, while levels of p21 and MDM2 are significantly lower in LFS MSCs.

(C) ChIP-PCR demonstrating lower p53 binding affinity to p21 and MDM2 promoter region in LFS MSCs. IgG ChIP is a negative control. Error bars indicate  $\pm$  SEM of triplicates.

(D) qRT-PCR for expression of p53 and its target genes after treatment of LFS and WT MSCs with the MDM2 inhibitor Nutlin-3 for 6 hr. Upregulation of the majority of p53 target genes is impaired in LFS MSCs in comparison with WT MSCs despite similar p53 expression in both MSC groups. qRT-PCR data are represented as mean  $\pm$  SEM; n = 3. See also Figure S2 and Table S1.



**Figure 3. LFS OBs Show Differentiation Defects and Oncogenic Properties**  
(A and B) Both AP (A) and alizarin red S (B) staining reveal the attenuation of OB differentiation in LFS OBs.  
(C and D) Expression of OB lineage markers (C) and transcriptional regulators (D) during the OB differentiation time course.

(legend continued on next page)



are largely composed of poorly differentiated or undifferentiated OBs (Tang et al., 2008), we asked if dysregulation of p53 signaling is responsible for the observed differentiation defects. LFS MSCs were induced to the OB lineage and the differentiation process was monitored over time. Several p53 targets were gradually induced in WT but not LFS OBs during differentiation (Figure S3A). Consistently, ChIP-PCR showed significantly reduced p53 binding to the p21 and MDM2 promoters in LFS OBs (Figure S3B). The decrease in p53 transcriptional activity in late stage osteogenic differentiation (day 17) was confirmed by a p53 reporter assay (Figure S3C). These findings suggest that p53 signaling is active in WT OBs but impaired in LFS OBs. AP staining for detecting bone-associated ALPL enzyme activity (Figure 3A) and alizarin red S staining reflective of mineral deposition by functional OBs (Figure 3B) showed slower differentiation in LFS MSCs. Mineral precipitations were observed on the surface of Petri dishes in WT but not LFS OBs (Figure S3D). Consistently, in comparison with WT OBs, LFS OBs showed lower expression of COL1A1 and ALPL (pre-OB markers), BGLAP/Osteocalcin, and PTH1R (mature OB markers), as well as FGF23 and MEPE (osteocyte markers) during osteogenesis (Figures 3C). Immunostaining confirmed that LFS OBs expressed lower BGLAP than WT OBs did (Figure S3E). Because osteogenic differentiation is controlled by several core transcriptional/epigenetic regulators, we monitored their expression levels during OB differentiation from LFS MSCs. Indeed, we found impaired upregulation of ZNF521 and ZEB1 (Figure 3D), indicating a defect in the normal OB gene regulatory network. Knockdown of p53 resulted in upregulation of osteogenic markers in LFS OBs and eliminated the osteogenic differentiation defect (Figure S3F) indicating that p53(G245D) may exert gain-of-function instead of loss-of-function effects in inhibiting osteogenic differentiation. Moreover, we noticed LFS OBs growing in randomly oriented piled-up foci rather than the two-dimensional monolayers of flattened cells (Figure S3G), a generally regarded manifestation of a transformed phenotype and an initiating step in tumorigenesis. To further investigate whether LFS OBs are able to recapitulate tumorigenic potential, we performed in vitro anchorage-independent growth (AIG) assays and in vivo xenografts. AIG assays showed clonal growth in soft agar by many LFS OBs (6 out of 9) but not in WT OBs, undifferentiated LFS or WT MSCs (Figure 3E). Performing xenografts in nude mice, we found tumorigenic ability in LFS OBs but not WT OBs (Figure 3F, right). The tumors demonstrated immature OB characteristics including AP activity, collagen matrix deposition but not mineralization (Figure 3F, left). The lack of in vitro and in vivo tumorigenic ability in LFS MSCs implies that OS may originate from immature or poorly differentiated OBs rather than MSCs (Figures 3E and 3F). To examine whether LFS OBs are able to gain malignancy during tumor progression, we performed serial

transplantation using an in ovo chick embryo chorioallantoic membrane (CAM) model. As shown in Figure S3H, the tumor sizes of LFS OBs increased in the second transplantation in comparison with the first transplantation. These results imply that similar to OS cells, LFS OBs contain a population of potential tumor-initiating cells (TICs) and these cells are enriched during in ovo primary transplantation and fuel secondary tumor growth. Interestingly, no additional increases in malignant tumor growth were seen following a tertiary transplantation. The persistence of a WT p53 allele during serial CAM transplantations (Figure S3I) provides a possible explanation for why there is no further gain of tumor growth in the tertiary transplantation. Taken together, these findings demonstrate that OS-related phenotypes (defective OB differentiation and tumorigenic ability) can be recapitulated in LFS iPSC-derived OBs.

### OS Spectrum Is Represented in LFS OBs

In order to gain insights into LFS-associated osteogenic defects and tumorigenesis, the global transcriptome was investigated by mRNA-seq during OB differentiation time courses (Table S2). Expression profiles of LFS and WT time course samples analyzed by Spearman's rank correlation demonstrated that gene expression profiles of LFS samples clustered together but were distinct from WT samples (Figure 4A). The non-negative matrix factorization (NMF) method for extracting relevant biological correlations based on gene expression data showed that at day 0 LFS and WT MSCs clustered together, while at days 7, 14, 17 differentiating LFS OBs are distinct from their WT counterparts (Figure 4B). These results suggest that WT and LFS MSC gene expression profiles are initially similar but diverge during subsequent OB differentiation. Alignment of reads at individual gene loci and quantification by fragments per kilobase of exon per million fragments mapped (FPKM) values confirmed the gradual increase of OB marker ALPL and skeletal development regulators HOXA10, IGF2, and CLEC3B in WT but not in LFS-derived cells (Figure S4A). Gene Ontology (GO) analyses using Network2Canvas further revealed that OB differentiation in WT MSCs (day 17 versus day 0) affects biological process genes mainly involved in skeletal system development and cell motility, whereas genes upregulated in LFS MSCs are primarily associated with an inflammatory response (Figure 4C, upper and middle). Expression levels of several skeletal system development-related genes were greatly increased during the WT osteogenesis time course but not in the LFS samples (Figure 4C, bottom). Moreover, expression levels of genes involved in positive regulation of cell differentiation and negative regulation of cell proliferation were significantly increased in WT OBs. In contrast, genes involved in positive regulation of cell cycle and mitosis were enriched in LFS OBs (Figure 4C, upper and middle). Using the Mouse Gene Atlas database, genes

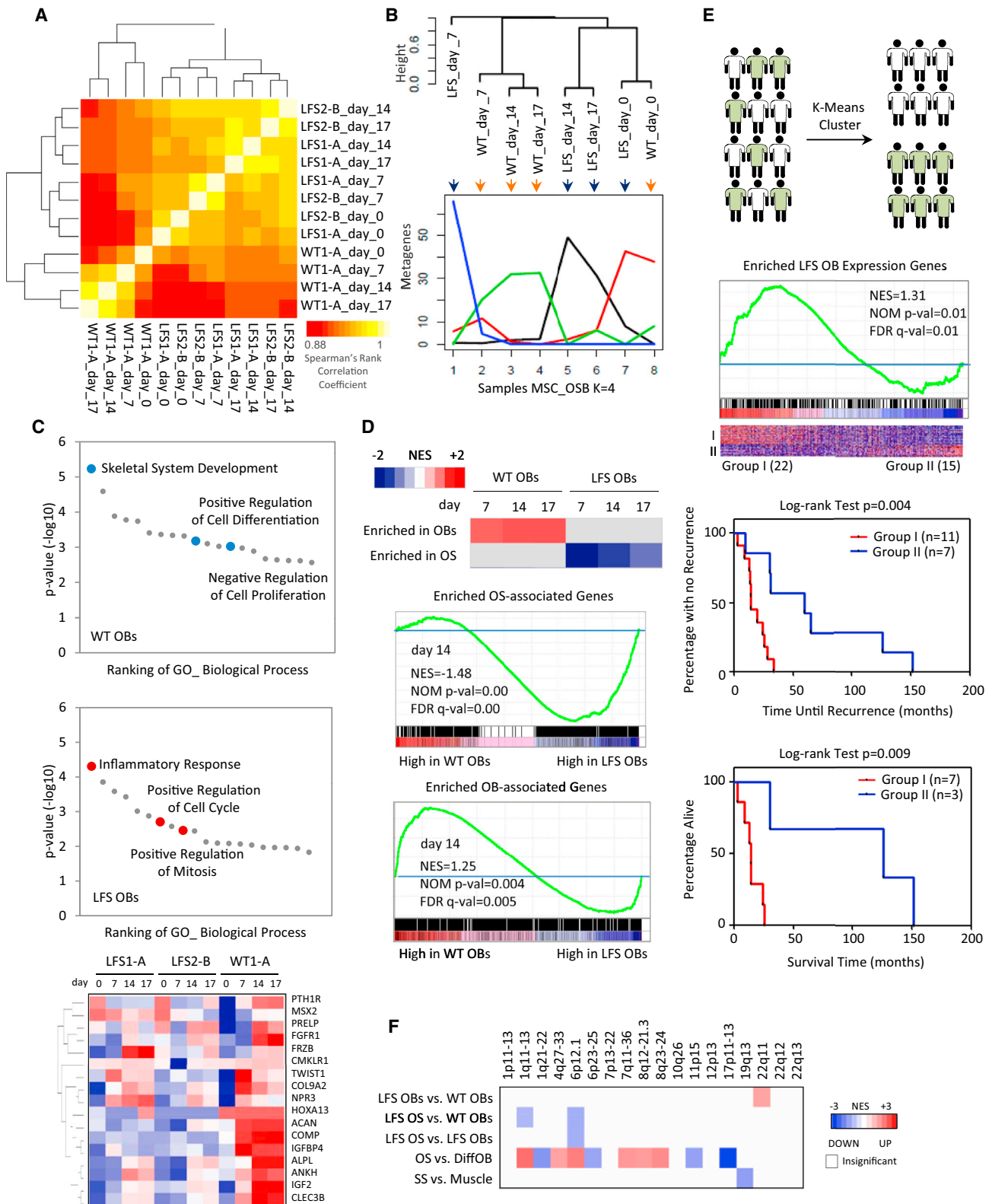
(E) In vitro AIG assay for tumorigenicity demonstrates colony numbers found in LFS OBs but not in WT OBs. Positive colonies after 1 month growth of differentiated OBs in either MSC or OB differentiation media are those larger than 50  $\mu\text{m}$  (scale bar, 50  $\mu\text{m}$ ).

(F) Tumor xenograft experiments by subcutaneous injection in *NU/NU* mice demonstrate that LFS OBs but not MSCs recapture in vivo tumorigenic ability. The LFS2-B OB-derived tumors were examined by H&E, AP, picrosirius red, and von Kossa stains to examine morphology, bone-associated AP expression, collagen production, and mineral deposits, respectively. Error bars represent  $\pm$  SEM. Scale bar, 1 cm.

(C–E) Error bars represent  $\pm$  SEM;  $n = 3$ .

See also Figure S3 and Table S1.





**Figure 4. Genome-Wide Transcriptome Analysis Reveals that LFS OBs Possess an OS Signature**

(A) Correlation matrix of LFS and WT osteogenic time course mRNA-seq results.

(B) Ordered tree linkage displaying sample clustering and metagenes representing the most variability associated with each differentiation transition.

(legend continued on next page)

upregulated at day 17 in WT OBs were more similar to the later-stage differentiated mouse OB gene profiles than were LFS OBs (Figure S4B). Further analyses of day 17 WT and LFS OB gene expression revealed that the expression pattern in WT OBs is similar to that of mouse OBs at day 21. In contrast, the gene expression pattern of LFS OBs is closest to that of mouse OBs at day 5 (Figure S4C). Consistent with our qRT-PCR analyses of p53 target gene expression (Figure S3A), gene set enrichment analysis (GSEA) confirmed that many known and predicted targets enriched in WT OBs relative to LFS OBs are found disproportionately in the set of significantly (>3-fold) upregulated genes in p53 transduced OS cells (Figure S4D). This supports the dysregulation of p53 function that occurs during LFS OB differentiation.

We next asked if the gene expression in LFS OBs is consistent with an oncogenic signature derived from OS cell lines. We identified both OB and OS signature genes by using GenePattern to compare the gene expression profiles between OB cells and OS lines (GSE39262). We then applied GSEA to compare enriched expressed genes between WT and LFS OBs against these signatures. As shown in Figure 4D, OS-associated genes are specifically enriched in LFS OBs; in contrast, OB-associated genes are specifically enriched in WT OBs. This finding strongly suggests that by the time of their generation, LFS OBs have already acquired OS characteristics. Furthermore, to determine whether the gene expression profiles of LFS OBs can potentially have prognostic value as measured by patient survival and tumor recurrence, we performed Kaplan-Meier analyses restricted to two separate subsets of patients, those who had or did not have an enriched LFS iPSC-derived OB gene expression signature as defined by enriched genes in LFS OBs versus WT OBs at day 17. We found that the enriched LFS OB-associated gene signature was significantly correlated with more rapid tumor recurrence and poorer survival ( $p = 0.004$  and  $p = 0.009$ , respectively) (Figure 4E). In summary, our LFS iPSC model not only recaptures the OS signature but can also predict clinical outcomes in patients.

Cytogenetic analyses of human OS have revealed numerous genomic alterations and rearrangements (Batanian et al., 2002; Bridge et al., 1993) that have been consistently replicated in a murine p53 conditional knockout OS model (Walkley et al., 2008). Since genomic alterations are common during cancer progression, it has been challenging to factor out their effects when attempting to interrogate the roles of tumor suppressor genes or oncogenes in tumor progression. To investigate if our LFS OBs provide a unique model to study early stages of tumor progression, we applied in silico cytogenetic region enrichment analysis (CREA) to our LFS OB samples to identify the potential presence of rearranged regions commonly found in human OS. The LFS OBs were compared with the gene expression signature of WT OBs at day 17 of OB differentiation,

while the synovial sarcoma (SS) was compared with normal muscle. As expected, OS cell lines showed significant enrichment at 10 of the 18 regions with known cytogenetic alterations in OS. In contrast, human SS were generally not associated with any alterations in these regions (Figure 4F), suggesting that these chromosomal regions are a specific feature of OS rather than a common feature of other cancers. Notably, in comparison with WT OBs, both LFS OBs and tumors showed negligible enrichment in these regions (1 and 2 out of 18 regions, respectively), implying that chromosomal rearrangements barely occur in these LFS OB-derived tumors (Figure 4F). These results demonstrate that LFS iPSC-derived OBs can serve as a useful system to study the early stages of OS progression caused solely by p53 mutation without interference by secondary genomic alterations.

### Impaired H19 Expression in LFS-Associated OS

Among 421 differentially expressed genes identified in comparisons between WT and LFS OBs, H19, highly expressed in OBs but not in bone marrow or osteoclasts (Figure S5A), warranted further in-depth analyses. Alignment and quantification of reads at the H19 locus and qRT-PCR showed H19 upregulation in WT but not LFS OBs during osteogenesis (Figures 5A and 5B). The low expression of H19 in LFS OBs was further confirmed in multiple LFS iPSC-derived OBs (Figure 5C). In comparison with bone/OB tissues and p53 WT cells, H19 expression is significantly decreased in OS and p53 mutant cells, respectively (Figures 5D and 5E). These findings suggested that H19 dysregulation is a common phenomenon in OS and is correlated with p53 status. RNAi-mediated knockdown of H19 in WT OBs led to decreased expression of osteogenic factors ZEB1 and ZNF521, pre-osteoblastic makers ALPL and COL1A1 as well as AP activity (Figure 5F). Supporting the positive regulatory role of H19 in osteogenesis, ectopic expression of H19 in LFS MSCs resulted in increasing osteogenic marker expression and reactivation of OBs with consequent mineral deposition (Figure 5G). Moreover, AIG and oncosphere assays demonstrated that in vitro tumorigenic activities of LFS OBs and OS TICs were suppressed by re-expressing H19 (Figures 5H and 5I). In LFS OBs assayed in ovo with the CAM assay and in vivo by the mouse tumor xenograft model, restoration of H19 expression not only reduced the incidence of tumor development but also decreased tumor size (Figures 5J and 5K). To further investigate if restoration of H19 has any therapeutic potential for OS treatment, H19 was transduced into OS cell lines, OSA and HOS. As shown in Figure S5B, H19 reduced the incidence of OS tumor development as well as tumor size, suggesting H19 as a therapeutic target. In comparison with the original OSA tumor, the H19-transduced OSA tumor demonstrated poorly differentiated osteoblastic characteristics including positive AP activity and collagen matrix deposition but not mineralization, implying that

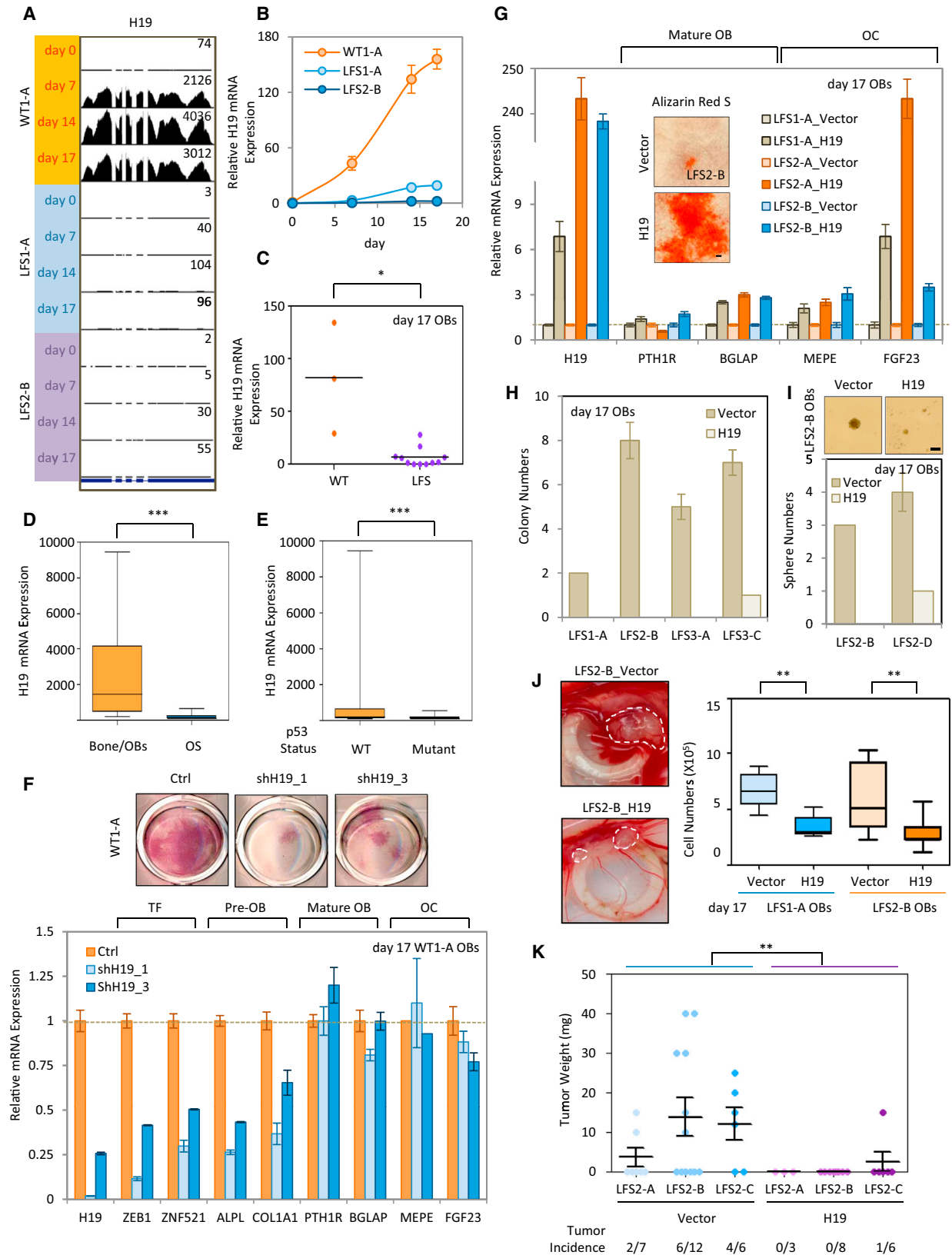
(C) GO biological processes associated with upregulated genes in WT and LFS OBs at day 17. FPKM values of skeletal system developmental genes are plotted as a heat map demonstrating their upregulation during WT but not LFS-derived osteogenic differentiation.

(D) GSEA indicates that OS-associated genes are enriched in LFS OBs while normal OB-associated genes are enriched in WT OBs.

(E) OS patients with the LFS OB signature show shorter tumor recurrence and poorer survival.

(F) CREA analysis reveals chromosomal integrity of LFS OBs and tumors engrafted in nude mice.

See also Figure S4 and Tables S2 and S6.



(legend on next page)

due to heterogeneity, a small portion of OS may gain of the ability to escape H19-induced OB differentiation and tumorigenic suppression. In summary, these findings emphasize the essential role of H19 in regulating osteogenesis and the potential of re-expressed H19 to rescue defective osteogenesis and suppress tumor growth in both LFS iPSC-derived OBs and OS cell lines.

Since H19 expression has been shown to be suppressed by p53 (Dugimont et al., 1998), we asked if this regulation was found in LFS patients with mutant p53. Consistent with the previous findings, activation of p53 by Nutlin-3 treatment decreased H19 expression by 20%–39% (Figure S5C) and RNAi-mediated p53 knockdown by two distinct RNAi molecules resulted in a 1.4- to 2.4-fold increase of H19 expression in WT MSCs (Figure S5D). In contrast, while suppression of H19 expression was not detected in Nutlin-3 treated LFS MSCs (Figure S5C), RNAi-mediated p53 knockdown led to a significant increase in H19 expression (7.1- to 19.3-fold) (Figure S5D), strongly suggesting that p53 mutants, at least p53(G245D), exert a gain-of-function effect in repressing H19. In agreement with the hypothesis that p53 does not bind the *H19* promoter region and that p53-mediated H19 repression occurs through other factors (Dugimont et al., 1998), p53 ChIP showed no enrichment of p53 binding in comparison with a control IgG pull-down (Figure S5E). To further explore whether other p53 mutants could promote this regulation, we transfected multiple variants of p53 into WT MSCs and examined their effects on H19 expression. As with p53(G245D), many p53 hotspot mutants (R175H, G245S, G248W, and R280K) exhibited stronger inhibition of H19 expression than did WT p53 (Figure S5F). This result demonstrates not only that the suppression of H19 expression by p53 mutants is common in LFS-associated OS but also that this is a general mechanism found in other LFS patients with distinct p53 mutations. Since it was shown that the tumorigenic ability of several p53 mutants is at least partially accounted for by their interaction with and inhibition of p63 and/or p73 functions (Di Como et al., 1999; Gaiddon et al., 2001), we investigated whether p53(G245D) could inhibit H19 expression directly via p63 and p73. Exogenous coimmunoprecipitation (coIP) showed that p53(G245D) and p53(R175H) but not p53(WT) can interact with p63 and p73 (Figure S5G). Depletion of p53(G245D) by siRNA increased p21mini promoter activity in two LFS-derived cells, LFS2-B ( $\Delta$ WT)-1 and LFS2-B( $\Delta$ WT)-2 MSCs that lacked a functional wild-type p53 allele due to insertion of a Neo<sup>R</sup> selection marker (D.-F. Lee and I.R. Lemischka, unpublished data).

Both p63 and p73 were able to activate the p21mini promoter to a greater extent upon p53(G245D) knockdown (Figure S5H), confirming the dominant-negative activity of p53(G245D) in regulating p63 and p73 function. However, ectopic expression of p63 and p73 does not alter H19 expression in WT MSCs (Figure S5I). These findings suggest that although one of the p53(G245D) gain-of-function effects in regulating LFS OS pathogenesis is through the suppression of normal p63/p73 function, this regulation is not involved in H19 transcriptional regulation.

Recent studies have found that p53 status may affect DNA methylation in the *H19* genomic locus in a clone-specific manner in iPSCs (Yi et al., 2012). To investigate whether impaired upregulation of H19 in LFS OBs is caused by hypermethylation on the *H19* locus, differentiating LFS OBs were treated with the demethylating agent 5'-aza-deoxycytidine (Decitabine). As shown in Figure S5J, H19 expression in LFS OBs was slightly increased upon Decitabine treatment but remained significantly lower than in WT OBs during OB differentiation, ruling out the possibility that lower H19 expression in LFS OBs is due to *H19* locus methylation. In fact, Decitabine-treated LFS OBs showed impaired OB differentiation ability (Figure S5K) and slightly increased in vitro tumorigenic potential (Figures S5L and S5M), suggesting that the clinical application of demethylating reagents, at least Decitabine, in treating LFS patients with OS is unlikely to provide much benefit.

### Involvement of Human IGN in Osteogenesis and Neurogenesis

Mouse H19 controls cell growth and development by regulating the expression of several imprinted genes within the IGN (Gabory et al., 2009). Interestingly, in comparison with other bone-associated tissues, the expression of mouse imprinted genes is enriched in differentiated OBs and many of these are themselves members of the IGN (9 out of 15; i.e., H19, Ndn, Igf2, Peg3, Zac1, Sgce, Dlk1, Mest, and Cdkn1c) (Figure S6A). Additionally, in comparison to normal mouse OBs, these imprinted genes, including H19, are significantly downregulated in mouse OS (Figure S6B). These results imply that the IGN may have a role in osteogenesis and that its dysregulation may promote OS in mice. We hypothesized that H19 suppresses LFS-associated OS through the imprinted gene regulatory system. Since a human IGN was not yet established, we first searched for genes frequently coexpressed with human imprinted genes and built a human IGN from a database of 79 human tissues (178 arrays)

### Figure 5. Involvement of H19 in LFS OB-Associated Defective OB Differentiation and Tumorigenesis

(A) Visualization of mRNA-seq short read mapping of H19 in UCSC Genome Browser.

(B) qRT-PCR shows impaired upregulated H19 in LFS during osteogenesis. Error bars indicate  $\pm$  SEM; n = 3.

(C) Multiple LFS OBs have impaired upregulation of H19 during OB differentiation.

(D and E) H19 expression is notably decreased in OS and p53 mutant cells. The analyses were performed using microarray data from GEO dataset GSE36001.

(F) RNAi-mediated knockdowns of H19 in WT MSCs leads to decreased expression of osteogenic genes as well as AP activity. qRT-PCR data are represented as mean  $\pm$  SEM; n = 3.

(G) Ectopic expression of H19 in LFS MSCs increases osteogenic gene expression and facilitates OB maturation. qRT-PCR data are represented as mean  $\pm$  SEM; n = 3. Scale bar, 100  $\mu$ m.

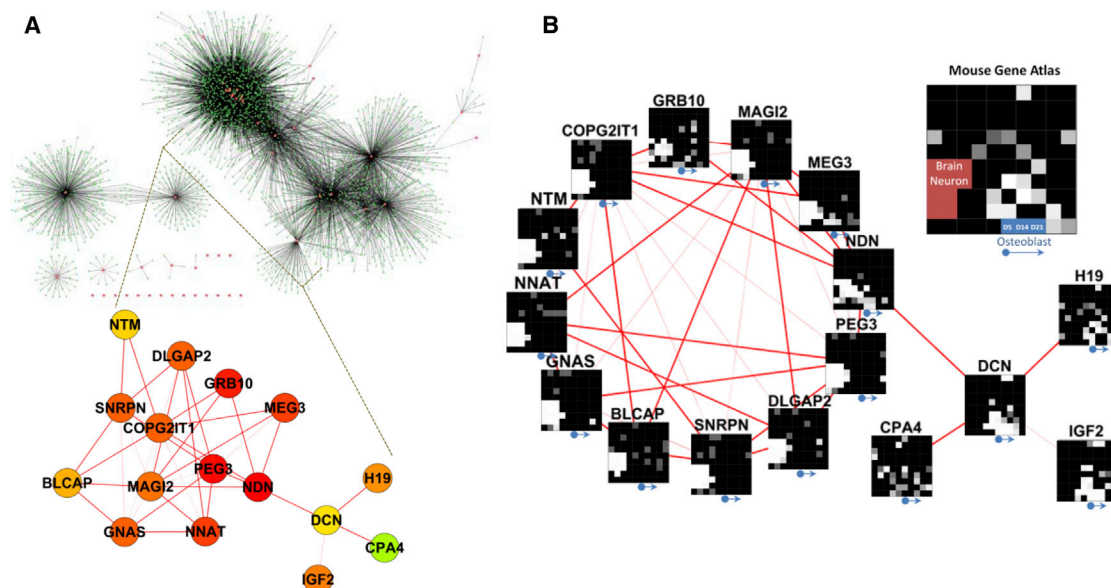
(H and I) AIG (H) and oncosphere (I) assays show repressed in vitro tumorigenic ability of LFS OBs upon restoration of H19. Error bars are  $\pm$  SEM; n = 3. Scale bar, 100  $\mu$ m.

(J) In ovo CAM assay indicates H19 suppresses tumorigenic ability of LFS OBs.

(K) In vivo tumor xenograft experiments indicating H19 suppression of tumorigenic ability of LFS OBs. Error bars represent  $\pm$  SEM.

See also Figure S5.





**Figure 6. A Network of Coregulated Human Imprinted Genes**

(A) Genes linked to human imprinted genes, including H19, identified from a set of 79 human tissues. Fifty-two imprinted genes are extracted from total 63 known and putative human imprinted genes. The main human IGN consists of 16 imprinted genes and contains two sub-networks.

(B) Network2Canvas analysis of these 16 imprinted coregulated genes by GO biological processes reveals that DCN-associated IGN primarily participates in osteogenesis and NDN-associated IGN in neurogenesis.

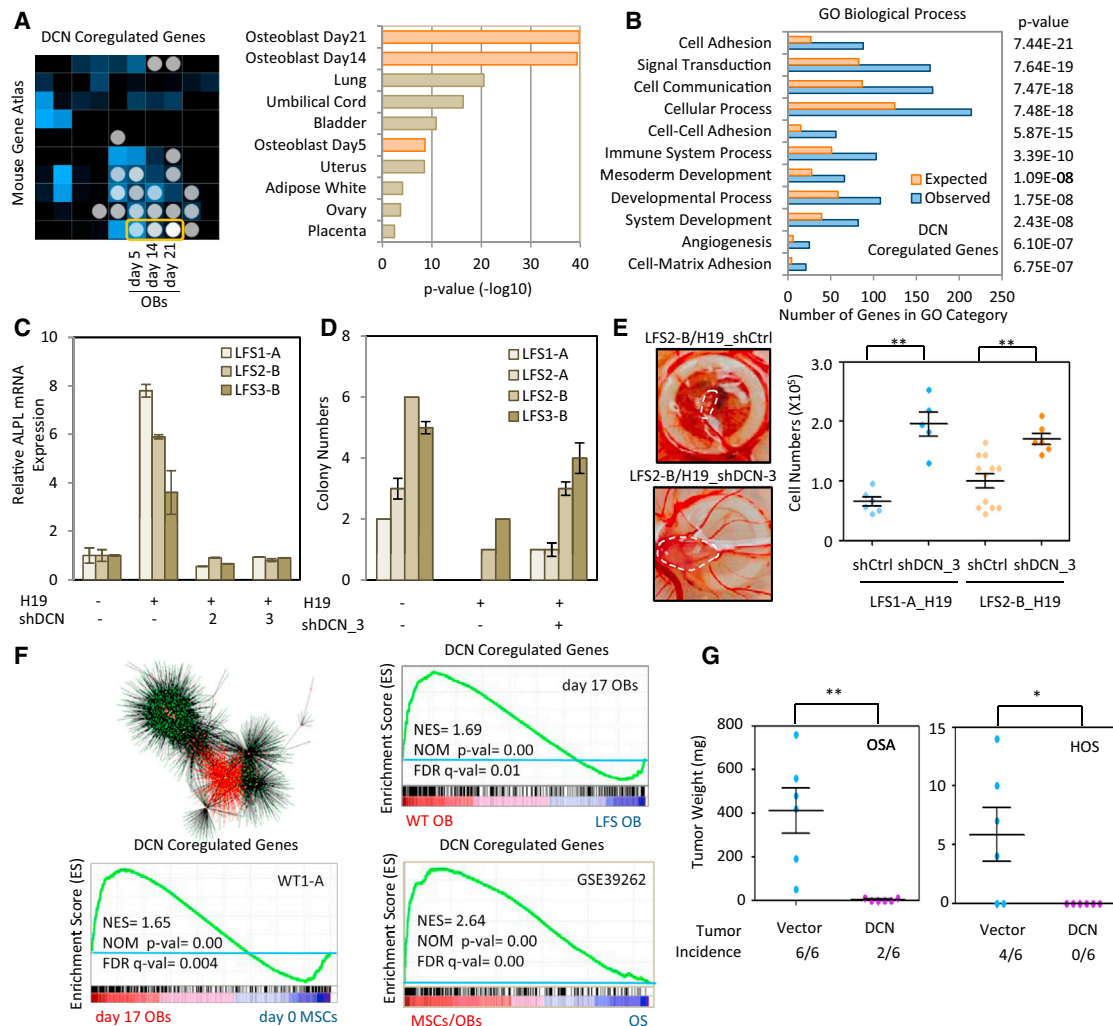
See also Figure S6 and Table S3.

(human U133A/GNF1H Gene Atlas; GSE1133). Fifty-two imprinted genes were extracted from total of 63 known and putative human imprinted genes (<http://www.geneimprint.com/>). The main human IGN is composed of 16 imprinted genes and divided into two sub-networks, the DCN and the NDN sub-IGN (Figure 6A). The DCN sub-IGN includes four imprinted genes (*H19*, *DCN*, *IGF2* and *CPA4*) and the NDN sub-IGN contains 12 imprinted genes (*NDN*, *PEG3*, *NNAT*, *MEG3*, *GNAS*, *MAGI2*, *COPG2IT1*, *GRB10*, *DLGAP2*, *SNRPN*, *NTM* and *BLCAP*). Tissue-specific gene expression defines their unique biological roles, and the enriched expressed genes, in general, have a role in maintaining tissue or cell-specific functions. Imprinted genes have been suggested to execute their functions through modulating the expression of their coregulated genes. Accordingly, we analyzed coregulated gene expression associated with these 16 imprinted genes (Table S3) using the Mouse Gene Atlas database. Interestingly, we found that the coregulated genes associated with individual imprinted genes in the DCN sub-IGN are primarily implicated in osteogenesis. In contrast, the coregulated genes associated with the NDN sub-IGN are mainly implicated in neurogenesis (Figure 6B). These findings strongly indicate that although individual imprinted genes participate in more diverse biological functions, the main roles of the entire IGN are in osteogenesis and neurogenesis. Gene expression analyses of 318 H19 coregulated genes suggest that these may function in early OB differentiation as well as lung, lactating mammary gland, and ciliary body development (Figure S6C). Interestingly, expression of these coregulated genes is mainly enriched in early but not late OB differentiation. This led us to suspect that H19 may function as an initiator of

osteogenesis and those downstream molecules, including coregulated imprinted genes, may function during later osteogenic stages.

#### DCN functions downstream of H19 in LFS-Associated OS Development

We next asked whether genes in the H19-associated IGN are responsible for the observed H19-modulated OB differentiation and oncogenic repression. We noticed that the imprinted gene *DCN*, encoding a matrix proteoglycan, is directly linked to H19 in the IGN (Figure 6B) and that *DCN*-coregulated genes are relatively overexpressed in OBs in a gradually increasing pattern during osteogenesis (Figure 7A). Because we also realized that the biological functions of *DCN*-coregulated genes are mainly in cell adhesion and the cell matrix (Figure 7B), both of which are known to regulate OB differentiation (Sosa-García et al., 2010), we pursued this candidate further. Alignment and quantification of reads at the *DCN* locus by FPKM values and qRT-PCR showed that *DCN* is gradually upregulated in WT but not LFS OBs during differentiation (Figures S7A and S7B). Decreased *DCN* expression in LFS OBs was further confirmed in OBs from multiple LFS iPSC lines (Figures S7C). *DCN* expression was significantly decreased in OS and p53 mutant cells in comparison with bone/OB tissues and p53 WT cells (Figures S7D and S7E). RNAi-mediated knockdown of *DCN* led to decreased expression of the osteogenic factor ZEB1, ALPL, and several osteogenic differentiation-associated genes (Figure S7F), supporting an essential role for *DCN* in normal osteogenesis. Furthermore, knockdown of H19 downregulated *DCN* while ectopic expression of H19 upregulated it in WT and LFS



**Figure 7. DCN Functions Downstream of H19 in LFS-Associated OS Development**

(A) GO biological processes examined by Network2Canvas indicate DCN coregulated genes are significantly upregulated during osteogenesis. (B) Panther analysis indicates DCN-coregulated genes are mainly involved in cell adhesion and cell matrix functions. (C) H19-mediated upregulation of pre-osteoblastic marker ALPL is abolished upon DCN depletion. Error bars indicate  $\pm$  SEM of triplicates. (D and E) RNAi-mediated DCN knockdown impairs H19-mediated inhibition of LFS OB tumorigenic activity in vitro (D) and in ovo (E). Error bars indicate  $\pm$  SEM,  $n = 3$  in (D). (F) GSEA of DCN-correlated gene expression indicates expression of DCN network genes during OB differentiation that is impaired in LFS OBs and OS cells. (G) DCN inhibits both OSA and HOS tumorigenesis. The sizes of HOS and OSA tumors were examined 1 and 2 months after subcutaneous injection, respectively. Error bars are  $\pm$  SEM;  $n = 6$ . See also Figure S7.

MSCs, respectively (Figure S7G). In agreement with these results, DCN expression was directly correlated with H19 expression in both WT and LFS samples as well as in OS cell lines (Figures S7H and S7I). In contrast, knockdown of DCN did not alter H19 expression (Figure S7J). These findings demonstrate that H19 functions as an upstream regulator of DCN expression. To further elucidate whether DCN functions as a downstream regulator involved in H19-mediated OB differentiation and tumor suppression, we depleted DCN in LFS OBs expressing ectopic H19 and found that H19-mediated upregulation of the pre-osteoblastic marker ALPL was abolished upon DCN knockdown

(Figure 7C) and that the in vitro and in ovo H19-mediated suppression of LFS OB tumorigenesis was inhibited (Figures 7D and 7E). These findings suggest that H19 regulation of osteogenesis and suppression of OS is, at least in part, mediated via DCN. In comparison with MSCs, GSEA of DCN-coregulated genes revealed their enrichment during OB differentiation (Figure 7F, left). Additionally, in comparison with WT OBs and MSC/OB tissues, DCN-coregulated genes were significantly decreased in LFS OBs and OS (Figure 7F, right), implying that DCN may negatively regulate OS development. Supporting its tumor suppressor function, ectopic expression of DCN reduced the incidence of

OSA and HOS-initiated tumor development as well as tumor size (Figure 7G). Taken together, these findings reveal that dysregulation of the H19-DCN IGN is strongly linked to LFS-associated osteogenic defects culminating in OS development.

## DISCUSSION

In vitro modeling of human disease has been greatly facilitated by iPSC methodologies (Takahashi et al., 2007; Takahashi and Yamanaka, 2006; Yu et al., 2007). Characterized by their ability to self-renew indefinitely and differentiate into all cell lineages of an organism, iPSCs provide a powerful system for human disease modeling. The p53 tumor suppressor is considered a promising therapeutic target to treat tumors with p53 mutations or deletions (Freed-Pastor and Prives, 2012). However, the lack of a reliable model limits the development of useful approaches to treat cancers caused by either genetic or somatic p53 mutations. Instead of regular application of clinical patient samples, cancer cell lines, and mouse models have been utilized to study p53 function. Here, we demonstrate the possibility of using LFS iPSCs to turn clinical samples into cell lines and models to study the pathological mechanisms caused by mutations in p53. This model system not only serves as an alternative tool to study p53 mutation-associated disorders but also provides substantial benefits for studying the role of p53 in the early stages of tumor development.

### The Role of Mutant p53 in Osteogenesis and OS

A series of studies have demonstrated that the p53 tumor suppressor promotes differentiation in a variety of cell types (Lee et al., 2012a; Molchadsky et al., 2010). Since some types of cancer, such as OS, are considered undifferentiated, it is logical to regard the cancer to be a defect in protective differentiation that would normally suppress unchecked cell proliferation and thus, prevent tumor development. However, recent in vivo evidence from p53 knockout and conditional MDM2 deletion mice, suggesting that wild-type p53 attenuates OB differentiation and bone development (Lengner et al., 2006; Wang et al., 2006), makes the situation far more complex. In our current studies, we found that H19 promotes OB differentiation and is repressed by p53; thus, providing a possible explanation for how p53 can suppress OB differentiation. Strikingly, the p53(G245D) mutant exerted a gain-of-function effect in repressing H19 transcription (Figures S5D and S5F), indicating that the defective OB differentiation in LFS OBs may result from inhibition of H19-mediated osteogenesis. In contrast to the p53(G245D) gain-of-function effect in downregulating H19 expression, this mutant exhibited a partial loss-of-function effect in upregulating the majority of known p53 downstream targets (e.g., MDM2, p21, and SFN). Advanced systems-level studies to characterize p53(G245D) function by identifying genome-wide differences between WT and mutant p53 will be needed to elucidate the comprehensive mechanisms involved in LFS-associated OS development.

### Is H19 a Tumor Suppressor or an Oncogene?

It has been suggested that H19 may act as a tumor suppressor (Hao et al., 1993; Yoshimizu et al., 2008). In contrast, several

in vitro culture experiments have suggested a controversial oncogenic role for H19 (Lustig-Yariv et al., 1997; Verkerk et al., 1997). This discrepancy could be explained by both differences in the experimental systems and the complexity of H19 functions in developmental and physiological processes. In our studies, H19 is both significantly upregulated during osteogenesis and commonly downregulated in OS, suggesting both differentiation-promoting and tumor-suppressing roles. Moreover, ectopic expression of H19 in LFS OBs restored normal osteogenic differentiation. H19 acts not only by directly modulating downstream targets as a lncRNA but also, indirectly controls an entire group of genes via its associated IGN. Notably, many distinct cancers are associated with dysregulation of imprinted genes (Joyce and Schofield, 1998). Such dysregulation may disrupt the H19-associated IGN and additional gene networks; thus, offering a possible explanation for the distinct effects, variously as an oncogene or a tumor suppressor observed after ectopic expression of H19. Moreover, it must also be acknowledged that the H19 locus may play more complex roles than regulation of the IGN, and these unidentified functions may also play roles in its bivalent function in tumorigenesis both in different tissues and at different tumor developmental stages.

### LFS iPSC Disease Model: An Alternative System to Study the Early Stages of OS Development

One key feature of clinically apparent OS is its numerous chromosomal alterations and rearrangements (Batanian et al., 2002; Bridge et al., 1993). A high level of genomic instability, in particular, chromosomal instability, is commonly found in OS. Tumor suppressor genes are frequently lost and oncogenes are duplicated. Because genomic instability is not only a consequence of tumor progression but also an active driver of tumor evolution, it creates a heterogeneous cell population and makes it more difficult to understand the initial steps of tumorigenesis. In comparison with normal differentiated OBs, human OS cell lines and a conditional mouse OS model (Walkley et al., 2008) showed strong enrichment of certain cytogenetic rearrangement regions. Because human/mouse OS lines are only isolated after many steps of tumor evolution, using these to study the initial stage of OS tumorigenesis is challenging. In marked contrast, LFS OBs and tumors showed a negligible degree of common OS cytogenetic rearrangements in comparison with WT OBs (Figure 4F), demonstrating the existence of a relatively intact genome. This relatively undisturbed genome also helps to explain the lower rate of and weak tumorigenicity of LFS OBs in vitro and in vivo. Thus we anticipate that LFS iPSC-derived OBs will be used in future studies focused on the role of mutant p53 in early OS progression prior to development of broad genomic alterations.

In summary, LFS iPSC-derived OBs not only provide a high-fidelity model system to elucidate the pathological mechanism of p53 mutant-associated OS development but also document a path for using LFS-associated gene expression patterns to predict clinical outcomes. More generally, iPSC approaches will also facilitate the definition of inherited versus somatically acquired causal components in many cancers. Further investigations to identify the regulatory mechanism of H19-DCN IGN and to develop drugs to activate H19 and DCN may have

powerful clinical implications for the treatment and/or prevention of OS in patients with either inherited or somatically acquired p53 mutations.

## EXPERIMENTAL PROCEDURES

### Somatic Cell Programming with Non-Integrating SeV

The fibroblasts of three LFS patients and two unaffected relatives were cultured and maintained in DMEM media supplemented 10% (vol/vol) Benchmark FBS (Gemini Bio-Product) and antibiotics. These fibroblasts were reprogrammed by transducing SeV expressing the four reprogramming factors OCT4, SOX2, KLF4, and c-MYC (CytoTune reprogramming kit, Invitrogen) according to manufacturer protocol. The reprogramming cells were maintained in hESC media (DMEM/F12 [Cellgro, Mediatech] containing 20% [vol/vol] KnockOut Serum Replacement [Invitrogen], L-glutamine, non-essential amino acids,  $\beta$ -mercaptoethanol, antibiotics and bFGF). After 3–4 weeks post-induction, individual clones with hESC/iPSC morphology and positive for TRA-1-60 and TRA-1-81 live staining were picked, passaged on MEFs and examined for loss of SeV by both staining with anti-SeV-specific antibody (PD029, MBL) and by qRT-PCR measurement of expression of the exogenous four factors. The specific qPCR primers targeting exogenous OCT4, SOX2, KLF4, and c-MYC are listed in Table S4.

### In Vitro Differentiation of iPSCs to MSCs

In vitro differentiation of WT and LFS iPSCs to MSCs was performed by a well-defined MSC differentiation protocol described previously (Lian et al., 2007). Briefly, iPSCs were seeded in gelatin-coated plates and cultured in MSC-differentiation media (DMEM supplemented with 10% Knockout serum replacement, 5 ng/ml FGF2 and 5 ng/ml PDGF-AB [PeproTech]) to induce differentiation. When differentiated cells were confluent, cells were trypsinized, split, and maintained. After 3 weeks of differentiation, the differentiated MSCs were sorted as the CD105 (eBioscience)-positive and CD24 (BD PharMingen)-negative cells by BD Arial in the Mount Sinai Flow Cytometry Shared Facility and expanded in MSC media (DMEM supplemented with 10% FBS). These differentiated MSCs were further examined for expression of other MSC surface markers CD44 (BD PharMingen), CD73 (BD PharMingen), and CD166 (BD PharMingen) as well as MSC-associated factors SNAI and VIM by immunostaining with anti-SNAI1 (Santa Cruz) and anti-VIM (Millipore) antibodies.

### In Vitro Osteogenic Differentiation of MSCs

LFS and WT iPSC-derived MSCs were plated in 12-well plate or 6-well plate at a density of  $1 \times 10^4$  cells or  $2 \times 10^4$  cells per well, respectively, in osteogenic differentiation medium ( $\alpha$ -MEM supplemented with 10% FBS, 0.1  $\mu$ M dexamethasone, 10 mM  $\beta$ -glycerol phosphate, and 200  $\mu$ M ascorbic acid) (Barberi et al., 2005). Cells were differentiated for specific time points as noted in the main text before characterization.

### In Vitro AIG and Oncosphere Assays

LFS and WT MSCs were cultured and passaged in 6-well plate at a density of  $2 \times 10^4$  cells per well. Cells were cultured in OB differentiation medium for 7 days, split and  $1 \times 10^4$  cells resuspended in OB differentiation medium with 0.4%–0.5% LMP agarose. The cell suspensions were then plated in 12-well plates containing solidified 0.8% agarose in OB differentiation medium. Cells were maintained in osteogenic differentiation medium for 1 month with medium changes every 3 days. Colony (considered to have a diameter  $\geq 50 \mu$ m) were counted under a microscope. For the oncosphere assay,  $2 \times 10^4$  7-day differentiated osteoblasts were washed by DPBS twice, resuspended in oncosphere medium ( $\alpha$ -MEM supplemented with 0.1  $\mu$ M dexamethasone, 10 mM  $\beta$ -glycerol phosphate, 200  $\mu$ M ascorbic acid, B27 supplement, 5  $\mu$ g/ml Heparin, 20 ng/ml bFGF, and 20 ng/ml EGF), and seeded in ultra-low attachment 6-well plates (3471, Corning). The number of oncospheres (diameter  $\geq 50 \mu$ m) was calculated after 12 days.

### Xenotransplantation

All animal procedures were performed in accordance with the Mount Sinai's Institutional Animal Care and Use Committee (IACUC).  $2 \times 10^6$  Matrigel-mixed

differentiated osteoblasts,  $2 \times 10^6$  Matrigel-mixed HOS (ATCC CRL-1543), and  $1 \times 10^6$  OSA (SJSA-1, ATCC CRL-2098) cells were injected subcutaneously into both right and left hind legs of 8-week-old immunocompromised nude mice (Charles River Laboratories). Tumors were excised around 6–10 weeks after injection. Tumors were weighed; fixed overnight in 10% neutral buffer formalin; embedded in paraffin; sectioned; and stained with H&E, AP, picrosirius red, and von Kossa stains to examine bone AP activity, AP expression, collagen matrix deposition, and mineralization, respectively, by HistoWiz.

### Statistical Analyses

Results are expressed as the mean and error bars represent SEM. Difference between two groups were examined by two-tailed unpaired t test. \*,  $p < 0.05$ ; \*\*,  $p < 0.01$ ; and \*\*\*,  $p < 0.001$ .

### ACCESSION NUMBERS

All mRNA-seq data are listed in Table S2 and deposited in NCBI-Gene Expression Omnibus database under accession number GSE58123.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.045>.

### ACKNOWLEDGMENTS

We thank P. Yang, X. Wei, C.-K. Chou, C.F. Pereira, Y. Liu, and X. Niu, for experimental assistance. We also thank E. Lopez-Rivera and M. Li for CAM assay assistance, the Mount Sinai Genomic Core Facility and Columbia Genome Center for mRNA-seq analysis, as well as Mount Sinai's Flow Cytometry Core for sorting MSCs. We gratefully acknowledge K.A. Moore, and A. Waghay for discussions. This research was funded by grants from the NIH (5R01GM078465), the Empire State Stem Cell Fund through New York State Department of Health (NYSTEM; C024176 and C024410) to I.R.L.; The University of Texas MD Anderson-China Medical University and Hospital Sister Institution Fund, The National Breast Cancer Foundation Inc., the Program for Stem Cell and Regenerative Medicine Frontier Research (NSC 102-2321-B-039-001 [Taiwan]) to M.-C.H.; the Canadian Breast Cancer Foundation-Prairies/NWT region to R.M.; and NIH Pathway to Independence Award (K99CA181496) to D.-F.L. B.C. is a trainee in NIDCR-Interdisciplinary Training in Systems and Developmental Biology and Birth Defects (T32HD075735).

Received: June 8, 2014

Revised: December 21, 2014

Accepted: February 9, 2015

Published: April 9, 2015

### REFERENCES

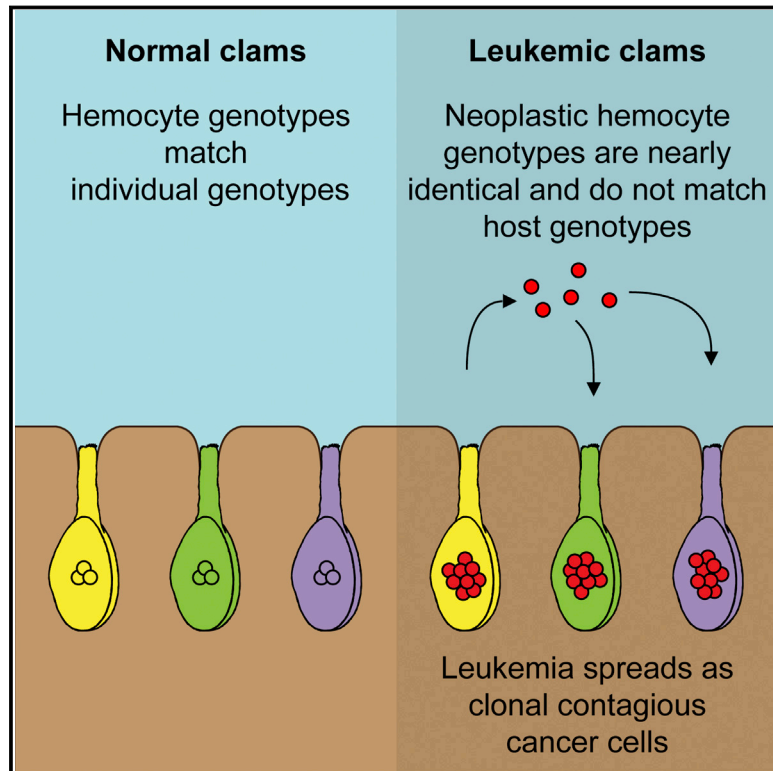
- Barberi, T., Willis, L.M., Socci, N.D., and Studer, L. (2005). Derivation of multipotent mesenchymal precursors from human embryonic stem cells. *PLoS Med.* 2, e161.
- Barley, R.D., Enns, L., Paterson, M.C., and Mirzayans, R. (1998). Aberrant p21WAF1-dependent growth arrest as the possible mechanism of abnormal resistance to ultraviolet light cytotoxicity in Li-Fraumeni syndrome fibroblast strains heterozygous for TP53 mutations. *Oncogene* 17, 533–543.
- Batanian, J.R., Cavalli, L.R., Aldosari, N.M., Ma, E., Sotelo-Avila, C., Ramos, M.B., Rone, J.D., Thorpe, C.M., and Haddad, B.R. (2002). Evaluation of paediatric osteosarcomas by classic cytogenetic and CGH analyses. *MP, Mol. Pathol.* 55, 389–393.
- Bridge, J.A., Bhatia, P.S., Anderson, J.R., and Neff, J.R. (1993). Biologic and clinical significance of cytogenetic and molecular cytogenetic abnormalities in benign and malignant cartilaginous lesions. *Cancer Genet. Cytogenet.* 69, 79–90.



- Choufani, S., Shuman, C., and Weksberg, R. (2010). Beckwith-Wiedemann syndrome. *Am. J. Med. Genet. C. Semin. Med. Genet.* 154C, 343–354.
- Di Como, C.J., Gaidon, C., and Prives, C. (1999). p73 function is inhibited by tumor-derived p53 mutants in mammalian cells. *Mol. Cell. Biol.* 19, 1438–1449.
- Dugimont, T., Montpellier, C., Adriaenssens, E., Lottin, S., Dumont, L., Iotsova, V., Lagrou, C., Stéhelin, D., Coll, J., and Cury, J.J. (1998). The H19 TATA-less promoter is efficiently repressed by wild-type tumor suppressor gene product p53. *Oncogene* 16, 2395–2401.
- Eggermann, T. (2010). Russell-Silver syndrome. *Am. J. Med. Genet. C. Semin. Med. Genet.* 154C, 355–364.
- Freed-Pastor, W.A., and Prives, C. (2012). Mutant p53: one name, many proteins. *Genes Dev.* 26, 1268–1286.
- Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* 85, 348–362.
- Gabory, A., Ripoche, M.A., Le Digarcher, A., Watrin, F., Ziyat, A., Forné, T., Jammes, H., Ainscough, J.F., Surani, M.A., Journot, L., and Dandolo, L. (2009). H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development* 136, 3413–3421.
- Gaidon, C., Lokshin, M., Ahn, J., Zhang, T., and Prives, C. (2001). A subset of tumor-derived mutant forms of p53 down-regulate p63 and p73 through a direct interaction with the p53 core domain. *Mol. Cell. Biol.* 21, 1874–1887.
- Hanel, W., Marchenko, N., Xu, S., Yu, S.X., Weng, W., and Moll, U. (2013). Two hot spot mutant p53 mouse models display differential gain of function in tumorigenesis. *Cell Death Differ.* 20, 898–909.
- Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E., and Tycko, B. (1993). Tumour-suppressor activity of H19 RNA. *Nature* 365, 764–767.
- Haydon, R.C., Luu, H.H., and He, T.C. (2007). Osteosarcoma and osteoblastic differentiation: a new perspective on oncogenesis. *Clin. Orthop. Relat. Res.* 454, 237–246.
- Joyce, J.A., and Schofield, P.N. (1998). Genomic imprinting and cancer. *MP, Mol. Pathol.* 51, 185–190.
- Lang, G.A., Iwakuma, T., Suh, Y.A., Liu, G., Rao, V.A., Parant, J.M., Valentin-Vega, Y.A., Terzian, T., Caldwell, L.C., Strong, L.C., et al. (2004). Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell* 119, 861–872.
- Lee, D.F., Su, J., Ang, Y.S., Carvajal-Vergara, X., Mulero-Navarro, S., Pereira, C.F., Gingold, J., Wang, H.L., Zhao, R., Sevilla, A., et al. (2012a). Regulation of embryonic and induced pluripotency by aurora kinase-p53 signaling. *Cell Stem Cell* 11, 179–194.
- Lengner, C.J., Steinman, H.A., Gagnon, J., Smith, T.W., Henderson, J.E., Kream, B.E., Stein, G.S., Lian, J.B., and Jones, S.N. (2006). Osteoblast differentiation and skeletal development are regulated by Mdm2-p53 signaling. *J. Cell Biol.* 172, 909–921.
- Li, F.P., and Fraumeni, J.F., Jr. (1969). Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome. *J. Natl. Cancer Inst.* 43, 1365–1373.
- Lian, Q., Lye, E., Suan Yeo, K., Khia Way Tan, E., Salto-Tellez, M., Liu, T.M., Palanisamy, N., El Oakley, R.M., Lee, E.H., Lim, B., and Lim, S.K. (2007). Derivation of clinically compliant MSCs from CD105+, CD24- differentiated human ESCs. *Stem Cells* 25, 425–436.
- Lustig-Yariv, O., Schulze, E., Komitowski, D., Erdmann, V., Schneider, T., de Groot, N., and Hochberg, A. (1997). The expression of the imprinted genes H19 and IGF-2 in choriocarcinoma cell lines. Is H19 a tumor suppressor gene? *Oncogene* 15, 169–177.
- Malkin, D., Li, F.P., Strong, L.C., Fraumeni, J.F., Jr., Nelson, C.E., Kim, D.H., Kassel, J., Gryka, M.A., Bischoff, F.Z., Tainsky, M.A., et al. (1990). Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250, 1233–1238.
- Mirzayans, R., Andrais, B., Scott, A., Paterson, M.C., and Murray, D. (2010). Single-cell analysis of p16(INK4a) and p21(WAF1) expression suggests distinct mechanisms of senescence in normal human and Li-Fraumeni Syndrome fibroblasts. *J. Cell. Physiol.* 223, 57–67.
- Molchadsky, A., Rivlin, N., Brosh, R., Rotter, V., and Sarig, R. (2010). p53 is balancing development, differentiation and de-differentiation to assure cancer prevention. *Carcinogenesis* 31, 1501–1508.
- Olive, K.P., Tuveson, D.A., Ruhe, Z.C., Yin, B., Willis, N.A., Bronson, R.T., Crowley, D., and Jacks, T. (2004). Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* 119, 847–860.
- Porter, D.E., Holden, S.T., Steel, C.M., Cohen, B.B., Wallace, M.R., and Reid, R. (1992). A significant proportion of patients with osteosarcoma may belong to Li-Fraumeni cancer families. *J. Bone Joint Surg. Br.* 74, 883–886.
- Sosa-García, B., Gunduz, V., Vázquez-Rivera, V., Cress, W.D., Wright, G., Bian, H., Hinds, P.W., and Santiago-Cardona, P.G. (2010). A role for the retinoblastoma protein as a regulator of mouse osteoblast cell adhesion: implications for osteogenesis and osteosarcoma formation. *PLoS ONE* 5, e13954.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861–872.
- Tang, N., Song, W.X., Luo, J., Haydon, R.C., and He, T.C. (2008). Osteosarcoma development and stem cell differentiation. *Clin. Orthop. Relat. Res.* 466, 2114–2130.
- Varley, J.M. (2003). Germline TP53 mutations and Li-Fraumeni syndrome. *Hum. Mutat.* 21, 313–320.
- Varrault, A., Gueydan, C., Delabre, A., Bellmann, A., Houssami, S., Aknin, C., Severac, D., Chotard, L., Kahli, M., Le Digarcher, A., et al. (2006). Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Dev. Cell* 11, 711–722.
- Verkerk, A.J., Ariel, I., Dekker, M.C., Schneider, T., van Gurp, R.J., de Groot, N., Gillis, A.J., Oosterhuis, J.W., Hochberg, A.A., and Looijenga, L.H. (1997). Unique expression patterns of H19 in human testicular cancers of different etiology. *Oncogene* 14, 95–107.
- Walkley, C.R., Qudsi, R., Sankaran, V.G., Perry, J.A., Gostissa, M., Roth, S.I., Rodda, S.J., Snay, E., Dunning, P., Fahey, F.H., et al. (2008). Conditional mouse osteosarcoma, dependent on p53 loss and potentiated by loss of Rb, mimics the human disease. *Genes Dev.* 22, 1662–1676.
- Wang, X., Kua, H.Y., Hu, Y., Guo, K., Zeng, Q., Wu, Q., Ng, H.H., Karsenty, G., de Crombrughe, B., Yeh, J., and Li, B. (2006). p53 functions as a negative regulator of osteoblastogenesis, osteoblast-dependent osteoclastogenesis, and bone remodeling. *J. Cell Biol.* 172, 115–125.
- Yi, L., Lu, C., Hu, W., Sun, Y., and Levine, A.J. (2012). Multiple roles of p53-related pathways in somatic cell reprogramming and stem cell differentiation. *Cancer Res.* 72, 5635–5645.
- Yoshimizu, T., Miroglio, A., Ripoche, M.A., Gabory, A., Vernucci, M., Riccio, A., Colnot, S., Godard, C., Terris, B., Jammes, H., and Dandolo, L. (2008). The H19 locus acts in vivo as a tumor suppressor. *Proc. Natl. Acad. Sci. USA* 105, 12417–12422.
- Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917–1920.

# Horizontal Transmission of Clonal Cancer Cells Causes Leukemia in Soft-Shell Clams

## Graphical Abstract



## Authors

Michael J. Metzger, Carol Reinisch, James Sherry, Stephen P. Goff

## Correspondence

spg1@columbia.edu

## In Brief

A fatal form of cancer is spreading between animals in the marine environment as a clonal horizontally transmissible cell, likely derived from a single original clam.

## Highlights

- Clam leukemia genotypes are distinct from their hosts and nearly identical to each other
- The transmissible cancer clone likely arose in a single individual
- Clam leukemia is transmitted horizontally between animals as contagious cancer cells
- Contagious cancer cell transmission may be widespread in the marine environment



Metzger et al., 2015, Cell 161, 255–263  
April 9, 2015 ©2015 Elsevier Inc.  
<http://dx.doi.org/10.1016/j.cell.2015.02.042>

# Horizontal Transmission of Clonal Cancer Cells Causes Leukemia in Soft-Shell Clams

Michael J. Metzger,<sup>1</sup> Carol Reinisch,<sup>2</sup> James Sherry,<sup>2</sup> and Stephen P. Goff<sup>1,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA

<sup>2</sup>Environment Canada, Water Science & Technology Directorate, Burlington, Ontario L7R 4A6, Canada

\*Correspondence: [spg1@columbia.edu](mailto:spg1@columbia.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.042>

## SUMMARY

Outbreaks of fatal leukemia-like cancers of marine bivalves throughout the world have led to massive population loss. The cause of the disease is unknown. We recently identified a retrotransposon, *Steamer*, that is highly expressed and amplified to high copy number in neoplastic cells of soft-shell clams (*Mya arenaria*). Through analysis of *Steamer* integration sites, mitochondrial DNA single-nucleotide polymorphisms (SNPs), and polymorphic microsatellite alleles, we show that the genotypes of neoplastic cells do not match those of the host animal. Instead, neoplastic cells from dispersed locations in New York, Maine, and Prince Edward Island (PEI), Canada, all have nearly identical genotypes that differ from those of the host. These results indicate that the cancer is spreading between animals in the marine environment as a clonal transmissible cell derived from a single original clam. Our findings suggest that horizontal transmission of cancer cells is more widespread in nature than previously supposed.

## INTRODUCTION

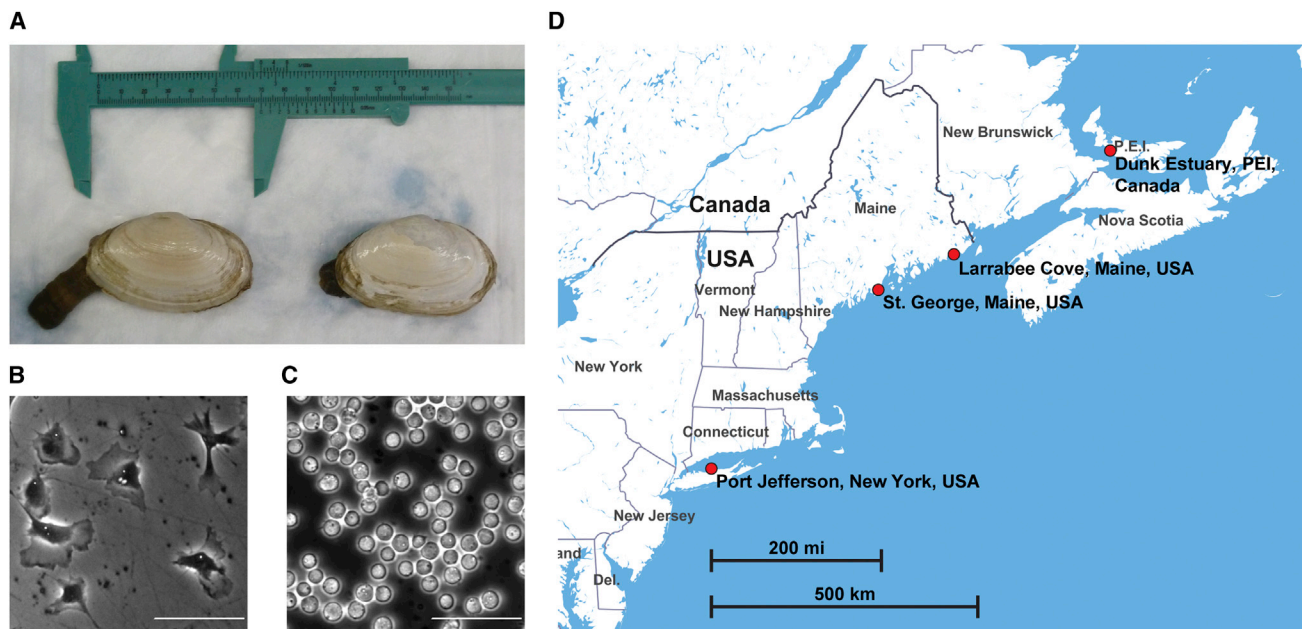
Cancers most often arise as a result of mutations accumulated in somatic cells during the lifetime of an organism and constitute a clonal expansion of a transformed cell with a genotype derived from the host. Tumors are generally not contagious or transmitted to other individuals, being subject to immune recognition and rejection based on polymorphic surface proteins, notably the major histocompatibility complex (MHC) in vertebrates. Some tumors are induced by infectious agents such as viruses, and though these agents can be contagious, each tumor still arises in the infected individual by transformation of somatic cells. Two cases are known in which a tumor cell itself naturally spreads among individuals as a contagious cell line. These are the canine transmissible venereal tumor (CTVT) (Murgia et al., 2006), transmitted by sexual contact, and the Tasmanian devil facial tumor disease (DFTD) (Pearse and Swift, 2006), transmitted between individuals by bites. In these two cases, the tumors exhibit a genotype that does not match that of their host. Instead, the tumor cells found in all of the affected animals

are a single clone with a unique genotype that reflects that of its original, primordial host.

Disseminated, or hemic, neoplasia is a leukemia-like cancer occurring in many marine bivalves, including clams, mussels, cockles, and oysters, and is characterized by amplification of cells in the hemolymph, the circulatory fluid of molluscs (Barber, 2004). The disease affects soft-shell clams (*Mya arenaria*) along the east coast of North America and was first documented in the species in the 1970s (Brown et al., 1978; Muttray et al., 2012; Yevich and Barszcz, 1978). The neoplastic hemocytes are characterized by distinctive morphology, new surface antigens, cytoplasmic mislocalization of the p53 tumor suppressor protein, loss of phagocytic abilities, greatly increased number (Figure 1), and dissemination of the cells into tissues (Barber, 2004; Miosky et al., 1989; Smolowitz et al., 1989; Walker et al., 2006, 2009). Most bivalve disseminated neoplasias are aneuploid, with higher than normal DNA content, and the neoplastic cells of *M. arenaria* are roughly tetraploid based on flow cytometric analysis of DNA content, although there is variation between individuals (Reno et al., 1994). The disease can be transmitted to naive animals by experimental transplantation of hemocytes (McLaughlin et al., 1992; Weinberg et al., 1997). A viral cause has been suspected, but no infectious agent has been confirmed (Taraska and Anne Böttger, 2013; Walker et al., 2009). The disease is ultimately fatal in the majority of affected clams and is contributing to depletion of the species in many areas along the east coast of North America (Barber, 2004; Cooper et al., 1982).

The detection of reverse transcriptase activity in neoplastic cells (House et al., 1998) suggested the possibility of retroelement involvement, and we used high-throughput sequencing of clam cDNA to identify a previously unknown LTR-retrotransposon in soft-shell clams, *Steamer*, whose expression was found to be strongly associated with disease (Arriagada et al., 2014). In normal clams, the genome contains 2–10 endogenous copies of the retrotransposon *Steamer* (normalized to a single copy gene), but in neoplastic hemocytes, the *Steamer* DNA copy number is massively amplified to 150–300 copies. The finding of several common *Steamer* integration sites in the neoplastic cells of multiple leukemic animals prompted further investigation of the genetics of this cancer.

Here, we first extend the observation that neoplastic cells contain common *Steamer* integration sites that are not present in normal animals or in normal tissues of diseased animals. These results have two possible explanations: either *Steamer* retrotransposons exhibit unprecedented selectivity for specific integration sites in these multiple neoplasms, or these



**Figure 1. Collection of Normal and Leukemic Soft-Shell Clams**

(A) Photograph of representative soft-shell clams (*M. arenaria*) with siphon partially extended.

(B) Hemolymph from a normal clam (NYTC-C6), showing attachment of hemocytes to the dish and extension of pseudopodia. Scale bar, 50  $\mu$ m.

(C) Hemolymph from a heavily leukemic clam (NYTC-C9) showing lack of attachment and rounded, refractile morphology. Hemolymph of the leukemic clam was diluted 1:100 to allow visualization of single cells. Scale bar, 50  $\mu$ m.

(D) Map of the eastern coast of North America with the locations of the clam collection sites (made with Mapbox Studio using data from OpenStreetMaps).

neoplasms did not arise independently but are descendants of a primordial leukemic cell carrying these common *Steamer* integrations, similar to the contagious cancers observed in dogs and Tasmanian devils. We therefore analyzed mitochondrial DNA (mtDNA) sequences and polymorphic microsatellite repeat loci and found that the genotypes of the neoplastic cells do not match those of their hosts. Furthermore, all neoplastic genotypes are nearly identical to each other, strongly arguing that soft-shell clam leukemia across the Atlantic coast is horizontally transmitted as a contagious cancer cell derived from a single clonal origin.

## RESULTS

### Common *Steamer* Integration Sites in Neoplastic Cells

Neoplastic cells of leukemic animals have a high copy number of the *Steamer* retrotransposon, but this high copy number is not present in normal cells. Analysis of DNA from different tissues of diseased animals showed that *Steamer* copy number was low in solid tissues and highest in hemocytes though with increases in *Steamer* copy number in some solid tissues, probably due to dissemination of neoplastic hemocytes (Figure 2). Preliminary sequence analysis of a small number of *Steamer* integration sites showed that neoplastic DNA contains new integration sites not present in normal clam DNA, and surprisingly, these new copies were often found at identical locations in different leukemic animals (Arriagada et al., 2014). We now extended this analysis to leukemic animals from independent populations

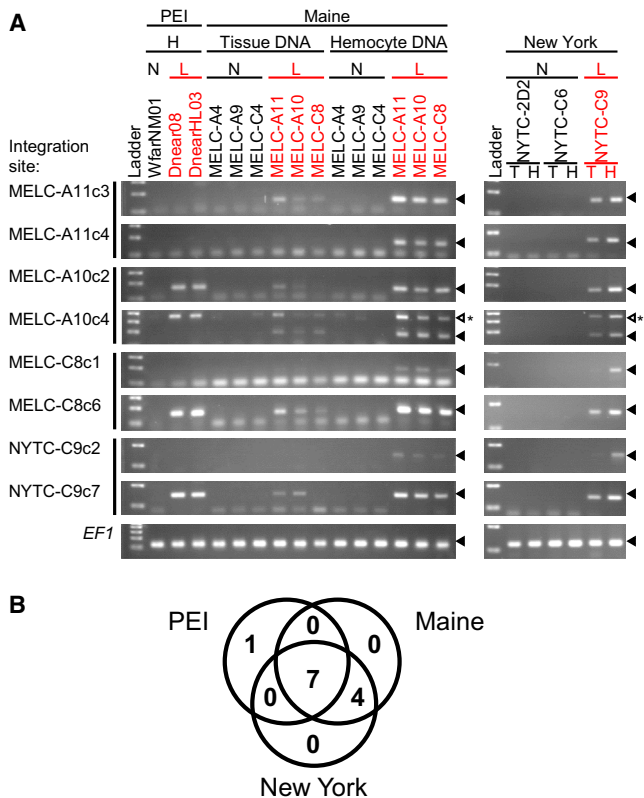
in New York, Maine, and Prince Edward Island (PEI), Canada. *Steamer* integration sites were cloned from multiple leukemic animals by inverse PCR, and diagnostic primers were designed to amplify each specific integration site. PCR tests using these primers then allowed us to score for the presence of each insertion in neoplastic and tissue DNA (Figure 3A). Of 12 integration sites tested, 7 were present in neoplastic DNA samples collected from all geographic locations, 4 more were common to the Maine and New York neoplasms, but not those from Canada, and 1 site was unique to neoplasms from Canada (Figure 3B). These cancer-specific integration sites were not found in any healthy animals and were only weakly detected in tissue DNA from leukemic animals, likely due to infiltration of the neoplastic cells into normal tissue.

### Common mtDNA SNPs in Neoplastic Cells

The observation of common integration sites raises the possibility of a clonal origin for all of the tumors. To test the hypothesis that soft-shell clam leukemia is horizontally transmitted between animals as a clonal transmissible cancer cell, we examined DNA from leukemic and normal clams for single-nucleotide polymorphisms (SNPs) in the sequences of mitochondrial genes encoding cytochrome c oxidase subunit I (COI) (Folmer et al., 1994) and cytochrome b (CYTB) (Table 1, Figure S1). DNA of all healthy animals ( $n = 11$ ) and the tissue DNA of all of the leukemic animals contained a common allele (C685) in the CYTB gene, while the neoplastic hemocyte DNA of all leukemic animals ( $n = 9$ ) carried a distinctive SNP (C685T). This observation supports the







**Figure 3. Clonal Steamer Integration Sites in Neoplastic Cells**

(A) To determine the presence of specific integration sites in different animals, inverse PCR products were cloned and sequenced from clams from three locations (two integration sites per animal). For each integration site, a reverse primer was designed to match the flanking genomic sequence. Diagnostic PCR using a common forward primer in the *Steamer* LTR and an integration site-specific reverse primer was used to determine the presence of each specific integration site in hemocyte (H) and tissue (T) DNA as indicated, of normal (N) and leukemic (L) clams. Sizes of the amplified DNAs were analyzed by agarose gel electrophoresis. Amplification of *EF1* is shown as a control. Filled triangles mark the position of migration of the amplicon. An open triangle marks an unexpected PCR product. This band is due to amplification of a second *Steamer* integration site present in neoplastic cells of all populations. (B) Venn diagram of the number of cancer-specific integration sites shared by neoplastic cells from the three locations (out of the 12 sites tested, including 4 sites cloned from PEI samples previously [Arriagada et al., 2014]). None of the 12 sites were present in any normal animal tested.

that the individual leukemias are not derived by independent oncogenic transformation of cells within each host but instead come from a single genetically unrelated parent. The data strongly argue that disseminated neoplasia is naturally spreading between animals as a transmissible cancer cell. Horizontal clonal transmission of cancer has only been observed in two other contagious cancers of mammals in the wild, the DFTD in Tasmanian devils (Pearse and Swift, 2006) and CTVT in dogs (Murgia et al., 2006).

It is remarkable that neoplastic cells from clam beds separated by hundreds of miles were found to be genetically nearly identical or very closely related. The mechanism by which the soft-shell leukemic cell line could be transferred from animal

to animal in the wild is not clear. In the two previously known cases of natural cancer cell transfer (DFTD and CTVT) (Murchison, 2008), physical contact is required for transmission of cancer cells to naive individuals (through biting or sexual contact, respectively), but adult clams are sessile and do not normally come in contact with one another. Clams do filter feed, however, raising the possibility that leukemia engraftment occurs through filtration of seawater contaminated with neoplastic cells. Each animal can filter several liters of seawater per hour, so very low concentrations of cells in seawater could be sufficient for transmission. One early study showed that hemocytes from a leukemic clam could survive in natural seawater conditions for >6 hr with minimal cell death, suggesting a plausible route for cancer cell transmission from one clam to another (Sunila and Farley, 1989). Direct surveys of seawater at the affected clam beds for tumor cells could provide support for this idea. It is currently unknown how neoplastic cells would be released by leukemic animals. Cells may be shed naturally during disease, expelled during spawning, or released after physical trauma, during predation, or at death of the leukemic individual. The stage of development when transmission might occur in nature is unknown, though experimental transmission of disease was found to be most efficient with mature animals of intermediate size (Taraska and Anne Böttger, 2013). It is possible that its transmission has been facilitated or accelerated by human intervention, as seed stocks have been transplanted between sites along the coast at several times in recent decades (Beal and Kraus, 2002).

The length of time since the original formation of the primordial soft-shell neoplasia is not yet clear. The disease has been documented in soft-shell clams since the 1970s (Brown et al., 1978; Yevich and Barszcz, 1978), and it is likely that all cases derive from the same clone, suggesting that it may be at least 40 years old and possibly much older. Its appearance must have been sufficiently far in the past to allow for its spread widely along the North Atlantic coast. Significant divergence was observed in cancer cell subgroups specific to the USA and Canada clam populations. Several small microsatellite expansions and deletions and at least one mtDNA SNP appear to have developed separately in the PEI subgroup of cancer cells. Further analysis of the origin and continued evolution of these cancer cell lineages may allow for estimation of the time of appearance of the original leukemia. The two other examples of transmissible tumors are of very different ages: the dog tumor is thought to have arisen 10,000–13,000 years ago (Murchison et al., 2014), while the Tasmanian devil tumor is of much more recent origin, perhaps arising only 20–30 years ago (Murchison, 2008).

It is plausible that the *Steamer* retrotransposon had a role in the development of the transmissible disseminated neoplasia. *Steamer* genomic copy number is greatly expanded in neoplastic cells (from 2–10 copies to 150–300 per haploid genome), and we show that the majority of the new integrations are common to all tested neoplastic cells and therefore likely occurred early in the evolution of the cancer. One or more of these common integration events may have directly caused initial oncogenic mutations, similar to the LINE1 integration near the *c-myc* gene in all known cases of CTVT (Katzir et al.,

**Table 1. Leukemia-Associated SNPs in *M. arenaria* COI and CYTB**

Location	Clam ID	Diagnosis	COI				CYTB							
			649		807		450		462		567		685	
			G217S		D269		P150		V154		F189		P229S	
			T	H	T	H	T	H	T	H	T	H	T	H
PEI	PEI-Dfar484	normal	–	G	–	T	–	T	–	T	–	C	–	C
PEI	PEI-Dfar490	normal	–	G	–	T	–	T	–	T	–	C	–	C
PEI	PEI-Dnear430	normal	–	G	–	T	–	T	–	T	–	C	–	C
PEI	PEI-Dnear432	normal	–	G	–	T	–	T	–	T	–	C	–	C
PEI	PEI-Dnear440	normal	–	G	–	T	–	T	–	T	–	C	–	C
PEI	PEI-Dnear446	normal	–	G	–	T	–	T	–	T	–	C	–	C
Maine	MELC-A4	normal	G	G	T	T	T	T	T	T	C	C	C	C
Maine	MELC-A9	normal	G	G	T	T	T	T	T	T	C	C	C	C
Maine	MELC-C4	normal	G	G	T	T	T	T	C	C	C	C	C	C
New York	NYTC-2D2	normal	G	G	T	T	T	T	T	T	C	C	C	C
New York	NYTC-C6	normal	G	G	T	T	T	T	T	T	C	C	C	C
PEI	PEI-Dfar488	leukemic	–	A	–	T	–	T	–	T	–	C	–	T
PEI	PEI-DnearHL03	leukemic	–	A	–	T	–	T	–	T	–	C	–	T
PEI	PEI-Dnear07	leukemic	G <sup>a</sup>	A	T	T	C <sup>a</sup>	T	T	T	C	C	C <sup>a</sup>	T
PEI	PEI-Dnear08	leukemic	G <sup>a</sup>	A	T	T	T	T	T	T	C	C	C <sup>a</sup>	T
Maine	MELC-A11	leukemic	G	G	C	T	T	T	T	T	C	C	C	T
Maine	MELC-A10	leukemic	G	G	T	T	T	T	T	T	C	C	C	T
Maine	MELC-C8	leukemic	G	G	T	T	T	T	T	T	C	C	C	T
Maine	MEHL01	leukemic	G	G	T	T	T	T	T	T	T	C	C	T
New York	NYTC-C9	leukemic	G	G	T	T	T	T	C <sup>a</sup>	T	C	C	C <sup>a</sup>	T

Nucleotide numbers based on the complete COI and CYTB genes in *M. arenaria* mtDNA genome (KJ755996). Boxes indicate discordance between hemocyte and tissue DNA. Three additional CYTB SNPs were identified: T442C (L148) in MELC-C4 and T658C (F220L) and T696C (I232) in MELC-A4. T, tissue; H, hemocyte; –, not done due to lack of availability of tissue samples.

<sup>a</sup>Hemocyte allele can also be observed in tissue DNA; see sequencing traces in Figures S1A and S1B.

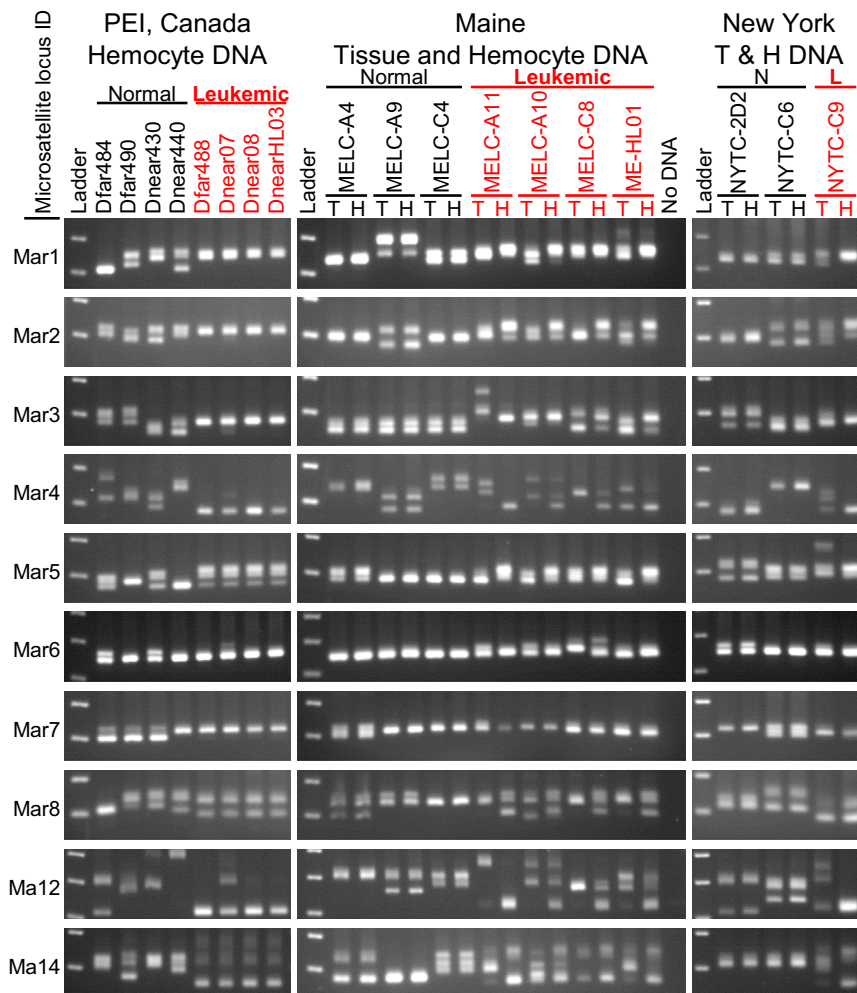
1985; Murgia et al., 2006), but alternatively they may represent passenger mutations that were fixed in the genome of the leukemia. Some of the integration sites are unique to the PEI or USA subgroups of neoplastic cells, and these may represent active gain of new copies by retrotransposition or loss of copies due to recombination or aneuploidy; these sites are unlikely to be a cause of the tumor. Further analysis of the genome and the expression profile of the neoplastic cells in comparison with those of normal hemocytes could provide clues to the cause of the transformed phenotype.

While all of the *M. arenaria* neoplastic cells that we have tested were derived from the same cell lineage, other leukemias may be derived from different clones or by more conventional mechanisms. Leukemia can be induced in clams by 5-bromodeoxyuridine (BrdU) injection (Oprandy and Chang, 1983; Taraska and Anne Böttger, 2013), and these cases must arise by de novo transformation of host cells and perhaps by induction of *Steamer* transposition. It has also been reported that clam leukemia could be transmitted between animals by a filterable agent (Oprandy et al., 1981; Taraska and Anne Böttger, 2013), but this is controversial (AboElkhair et al., 2012; McLaughlin et al., 1992).

Transmissible disseminated neoplasias have been reported in other bivalve species, including other clams, oysters, mussels,

and cockles (Barber, 2004), and these diseases may also be caused by a clonal transmissible cancer cell. A recent report (Vassilenko et al., 2010) that neoplastic hemocytes in mussels (*Mytilus trossulus*) on the west coast of North America share a common set of synonymous SNPs in the *p53* gene is consistent with an independent development of a transmissible clonal cancer in that species.

Flow cytometric analyses of DNA content of disseminated neoplasias in many bivalve species have identified characteristic ploidy levels—indeed, polyploidy has been used as diagnostic for disease (Delaporte et al., 2008). In *M. arenaria* specifically, the observation of roughly tetraploid DNA content in neoplastic cells (Delaporte et al., 2008; Reno et al., 1994) is consistent with a clonal contagious cancer and with the observation of up to four unique microsatellite alleles at a single locus in neoplastic cells. Given the current findings, previous observations of abnormal ploidy in bivalve disseminated neoplasia suggest that many of these diseases in other species may represent independent contagious cancer lineages. For example, mussel-disseminated neoplasia has been reported to come in two types, tetraploid or pentaploid (Moore et al., 1991), consistent with either two independent cancer lineages or one that has evolved into subgroups with divergent ploidy, although a more recent study found a wider distribution of



**Figure 4. Amplification of Microsatellite Loci from Tissue and Hemocyte DNA from Normal and Leukemic Clams**

PCR products using primers flanking ten microsatellite loci in hemocyte DNA and tissue DNA were displayed by electrophoresis on 2.5% agarose gels and visualized by staining. Different alleles are determined by the sizes of the amplicons, with one band observed for animals homozygous at a particular locus and two or more for heterozygotes and polyploid neoplastic cells. Each row of gels represents amplification from a single microsatellite locus, as labeled on the left. Hemocyte DNA is shown for normal and leukemic clams from PEI, Canada, and both tissue (T) and hemocyte (H) DNA is shown for Maine and New York clams. These data show that the leukemic hemocyte microsatellite alleles are identical to each other and distinct from their host tissue.

regulate MHC expression, and the low genetic diversity of devils limits the ability of hosts to recognize DFTD as foreign (Siddle and Kaufman, 2015; Siddle et al., 2013). Contagious cancer is a serious threat to marine invertebrates, leading to severe mortalities during outbreaks in soft-shell clams and possibly leading to mass mortalities in many other species. The disease represents a significant selective pressure and supports the hypothesis that histocompatibility could have evolved, in part, due to selective pressure to prevent malignancy (Murgia et al., 2006) rather than simply being a secondary consequence of pressure by infec-

aneuploidy between individuals (Vassilenko and Baldwin, 2014). Multiple abnormal ploidy levels have also been observed in disseminated neoplasia in cockles (*Cerastoderma edule*) (da Silva et al., 2005). We are currently investigating the hypothesis suggested by the current findings that disseminated neoplasia in other bivalve species represent independent lineages of contagious cancer. Preliminary mtDNA sequencing of cockles with disseminated neoplasia suggests a cockle-derived contagious cancer line (data not shown).

It is possible that the disseminated neoplasia of *M. arenaria* could transmit to other species, though there is as of yet no evidence for this. There may be mechanisms by which a host animal can reject colonization by neoplastic cells of another species. Molluscs are not known to have a self/nonself recognition system similar to the MHC system of vertebrates, and their lack of MHC may make them highly susceptible to this form of infectious malignancy. Indeed, in the cases of CTVT and DFTD, there are unique mechanisms by which the cancers avoid MHC recognition. In CTVT, expression of MHC I and II is downregulated during active growth of the tumor, and their expression leads to immune recognition and clearance of the tumor in most dogs (Yang et al., 1987). DFTD cells also down-

regulate MHC expression, and the low genetic diversity of devils limits the ability of hosts to recognize DFTD as foreign (Siddle and Kaufman, 2015; Siddle et al., 2013). Contagious cancer is a serious threat to marine invertebrates, leading to severe mortalities during outbreaks in soft-shell clams and possibly leading to mass mortalities in many other species. The disease represents a significant selective pressure and supports the hypothesis that histocompatibility could have evolved, in part, due to selective pressure to prevent malignancy (Murgia et al., 2006) rather than simply being a secondary consequence of pressure by infec-

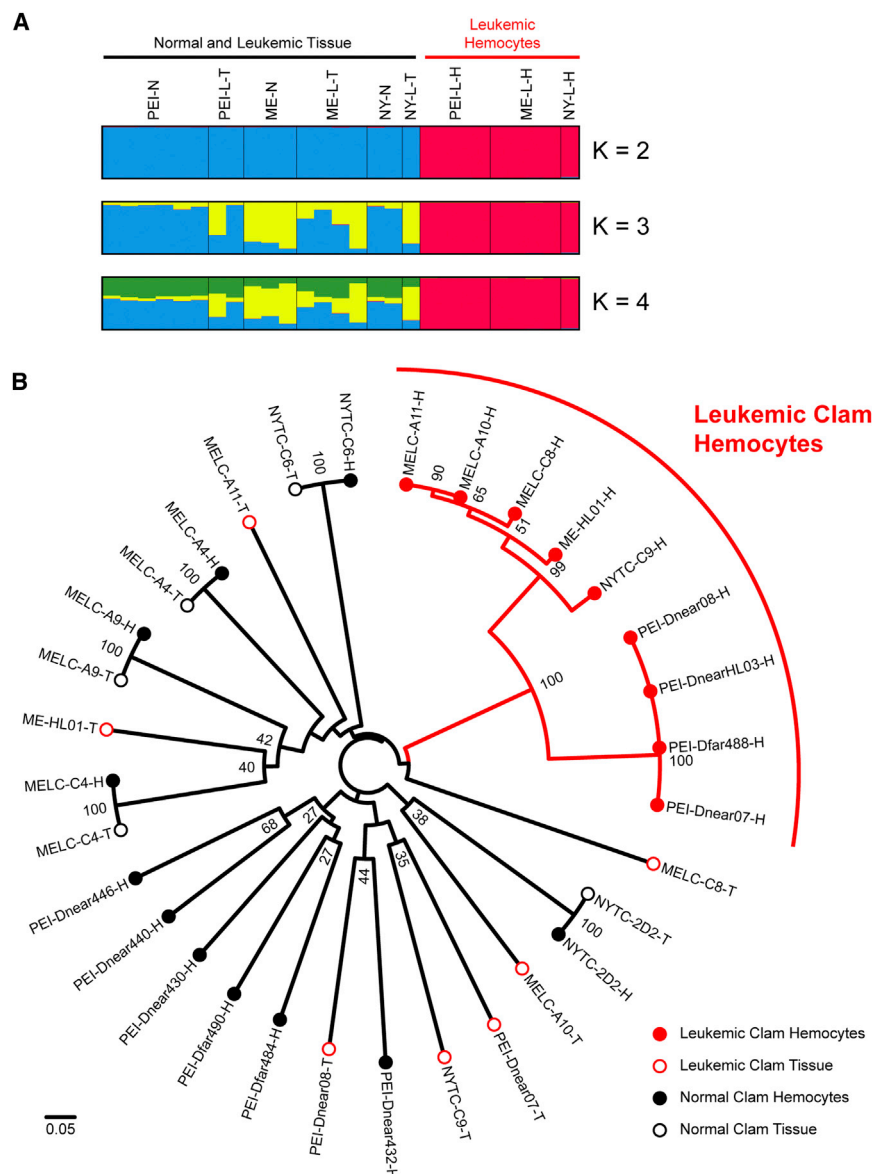
Natural horizontal transmission of cancer between individuals has been considered a rare phenomenon, restricted to two exceptional cases in mammals. Our finding of the horizontal transmission of a clonal clam leukemia extends the phenomenon to the marine environment and demonstrates that this mechanism is more widespread in nature than previously supposed.

## EXPERIMENTAL PROCEDURES

### Sample Collection

Soft-shell clams (*M. arenaria*) were collected from several locations (Figure 1). Clams from the Dunk estuary on Prince Edward Island (PEI), Canada were collected and diagnosed as described previously (Arriagada et al., 2014). Clams from Larrabee Cove, Maine (MELC) were collected by Brian Beal





**Figure 5. Analysis of Microsatellite Alleles**

(A) CLUMPAK display of STRUCTURE (Pritchard et al., 2000) analysis of ten microsatellite loci showing the major population clustering with varying number of clusters (K) from K = 2–4. Each vertical bar represents a normal clam genotype or either tissue or hemocyte genotypes from leukemic clams with the colors representing cluster identity.

(B) Neighbor-joining tree constructed with genetic distances based on ten microsatellite loci using Bruvo's method for analysis of loci from individuals with variable ploidy (Bruvo et al., 2004) calculated using the *poppr* package with R (Kamvar et al., 2014). Bootstrap values above 25 are shown at nodes. The scale bar represents a genetic distance of 0.05 (where 0 represents completely identical genotypes and 1 represents no common alleles). Each sample is marked as tissue (T, closed circle) or hemocytes (H, open circle) of normal (black) or leukemic (red) clams. The leukemic genotypes all cluster together in two branches clearly apart from the normal genotypes. Allele sizes are listed in Table S2, and further information is available in the Extended Experimental Procedures.

#### Microsatellite Analysis

STRUCTURE (Pritchard et al., 2000) was used to determine the population clustering of the clam tissue and hemocyte DNA. As up to four unique alleles were detected in leukemic DNA, loci were considered to be tetraploid, and clams with lower ploidy were considered to have missing data for the additional alleles, which are ignored in analysis. Cluster values of K = 1 to 10 were used with an admixture model with a burnin of 50,000 and 50,000 replicates after burnin. CLUMPAK (Kopelman et al., 2015) was used to graphically analyze the major clusters across ten STRUCTURE runs. Data were also recoded and reanalyzed as binary absent/present alleles (Rodzen and May, 2002), with nearly identical results (data not shown).

Genetic distance was calculated using Bruvo's band-sharing method assuming infinite alleles (Bruvo et al., 2004), as it was created specifically

(University of Maine at Machias) as described previously, and one additional leukemic clam (ME-HL01) was collected by a commercial source near St. George, Maine with the help of Charles Walker (University of New Hampshire). Clams from New York (NYTC) were acquired from a commercial market and collected from the north shore of Long Island, New York near Port Jefferson. Individual clam IDs reflect the location and the ID number assigned at the time of collection.

Diagnosis of clams from New York and Maine was conducted by extracting hemolymph from the pericardial region using a 26 gauge needle. Six to eight drops of hemolymph were placed in a well of a 96-well plate and left undisturbed at 10°C for 1 hr before morphological analysis using phase-contrast microscopy. Normal clam hemocytes attached to the dish (Figure 1A), and clams with >50% rounded, refractile cells were considered leukemic (Figure 1B).

#### PCR and qPCR

Primers and conditions used for qPCR, diagnostic PCR for specific integration sites, *COI* (Folmer et al., 1994) and *CYTb* PCR, and PCR of microsatellite loci are provided in Extended Experimental Procedures and Table S1.

for comparisons of individuals with different ploidy levels, and a neighbor-joining tree was constructed (with 100 bootstraps) using the *poppr* package for R (Kamvar et al., 2014) and displayed using Figtree. Alternate analyses using the genome addition and combined genome addition/genome loss models generated nearly identical results (data not shown). Further details of the microsatellite analysis can be found in the Extended Experimental Procedures.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, one figure, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.042>.

#### AUTHOR CONTRIBUTIONS

S.P.G. and M.J.M. planned the experiments and wrote the paper. M.J.M. conducted the experiments. C.R. and J.S. collected and diagnosed clams from PEI.

## ACKNOWLEDGMENTS

We thank Gloria Arriagada for work on the *Steamer* element, Brian Beal (University of Maine at Machias) for *M. arenaria* samples from Maine, Annette Murray for help in collection and diagnosis of *M. arenaria* from PEI, and Charles Walker and Sarah Mercier (University of New Hampshire) for help in acquisition of leukemic sample ME-HL01. We also thank Anne Gifford for putting the laboratories in contact to initiate the studies. This work was supported by the Howard Hughes Medical Institute and by the National Institutes of Health (NIH) Training Grant T32 CA009503 (to M.J.M.).

Received: November 29, 2014

Revised: January 15, 2015

Accepted: January 27, 2015

Published: April 9, 2015

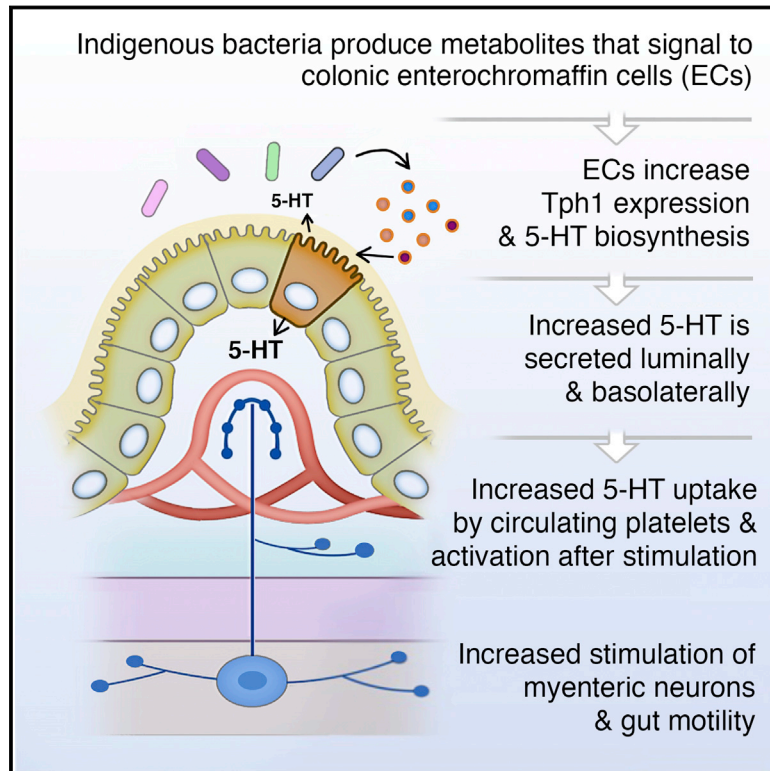
## REFERENCES

- AboElkhair, M., Iwamoto, T., Clark, K.F., McKenna, P., Siah, A., Greenwood, S.J., Berthe, F.C., Casey, J.W., and Cepica, A. (2012). Lack of detection of a putative retrovirus associated with haemic neoplasia in the soft shell clam *Mya arenaria*. *J. Invertebr. Pathol.* 109, 97–104.
- Arriagada, G., Metzger, M.J., Muttray, A.F., Sherry, J., Reinisch, C., Street, C., Lipkin, W.I., and Goff, S.P. (2014). Activation of transcription and retrotransposition of a novel retroelement, *Steamer*, in neoplastic hemocytes of the mollusk *Mya arenaria*. *Proc. Natl. Acad. Sci. USA* 111, 14175–14180.
- Barber, B.J. (2004). Neoplastic diseases of commercially important marine bivalves. *Aquat. Living Resour.* 17, 449–466.
- Beal, B.F., and Kraus, M.G. (2002). Interactive effects of initial size, stocking density, and type of predator deterrent netting on survival and growth of cultured juveniles of the soft-shell clam, *Mya arenaria* L., in eastern Maine. *Aquaculture* 208, 81–111.
- Brown, R.S., Wolke, R.E., Salla, S.B., and Brown, C.W. (1978). Prevalence of neoplasia in 10 New England populations of the soft-shell clam (*Mya arenaria*). *Ann. N.Y. Acad. Sci.* 298, 522–534.
- Bruvo, R., Michiels, N.K., D'Souza, T.G., and Schulenburg, H. (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Mol. Ecol.* 13, 2101–2106.
- Cooper, K.R., Brown, R.S., and Chang, P.W. (1982). The Course and Mortality of a Hematopoietic Neoplasm in the Soft-Shell Clam, *Mya-Arenaria*. *J. Invertebr. Pathol.* 39, 149–157.
- da Silva, P.M., Soudant, P., Carballal, M.J., Lambert, C., and Villalba, A. (2005). Flow cytometric DNA content analysis of neoplastic cells in haemolymph of the cockle *Cerastoderma edule*. *Dis. Aquat. Organ.* 67, 133–139.
- De Tomaso, A.W., Nyholm, S.V., Palmeri, K.J., Ishizuka, K.J., Ludington, W.B., Mitchel, K., and Weissman, I.L. (2005). Isolation and characterization of a protochordate histocompatibility locus. *Nature* 438, 454–459.
- Delaporte, M., Synard, S., Pariseau, J., McKenna, P., Tremblay, R., Davidson, J., and Berthe, F.C. (2008). Assessment of haemic neoplasia in different soft shell clam *Mya arenaria* populations from eastern Canada by flow cytometry. *J. Invertebr. Pathol.* 98, 190–197.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* 3, 294–299.
- House, M.L., Kim, C.H., and Reno, P.W. (1998). Soft shell clams *Mya arenaria* with disseminated neoplasia demonstrate reverse transcriptase activity. *Dis. Aquat. Organ.* 34, 187–192.
- Kamvar, Z.N., Tabima, J.F., and Grünwald, N.J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281.
- Katzir, N., Rechavi, G., Cohen, J.B., Unger, T., Simoni, F., Segal, S., Cohen, D., and Givol, D. (1985). "Retroposon" insertion into the cellular oncogene c-myc in canine transmissible venereal tumor. *Proc. Natl. Acad. Sci. USA* 82, 1054–1058.
- Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* Published online February 27, 2015. <http://dx.doi.org/10.1111/1755-0998.12387>.
- Krapal, A.M., Popa, O.P., Iorgu, E.I., Costache, M., and Popa, L.O. (2012). Isolation and characterization of new microsatellite markers for the invasive softshell clam, *Mya arenaria* (L.) (Bivalvia: Myidae). *Int. J. Mol. Sci.* 13, 2515–2520.
- McLaughlin, S.M., Farley, C.A., and Hetrick, F.M. (1992). Transmission studies of sarcoma in the soft-shell clam, *Mya arenaria*. *In Vivo* 6, 367–370.
- Miosky, D.L., Smolowitz, R.M., and Reinisch, C.L. (1989). Leukemia cell specific protein of the bivalve mollusc *Mya arenaria*. *J. Invertebr. Pathol.* 53, 32–40.
- Moore, J.D., Elston, R.A., Drum, A.S., and Wilkinson, M.T. (1991). Alternate pathogenesis of systemic neoplasia in the bivalve mollusc *Mytilus*. *J. Invertebr. Pathol.* 58, 231–243.
- Murchison, E.P. (2008). Clonally transmissible cancers in dogs and Tasmanian devils. *Oncogene* 27 (2), S19–S30.
- Murchison, E.P., Wedge, D.C., Alexandrov, L.B., Fu, B., Martincorena, I., Ning, Z., Tubio, J.M., Werner, E.I., Allen, J., De Nardi, A.B., et al. (2014). Transmissible [corrected] dog cancer genome reveals the origin and history of an ancient cell lineage. *Science* 343, 437–440.
- Murgia, C., Pritchard, J.K., Kim, S.Y., Fassati, A., and Weiss, R.A. (2006). Clonal origin and evolution of a transmissible cancer. *Cell* 126, 477–487.
- Muttray, A., Reinisch, C., Miller, J., Ernst, W., Gillis, P., Losier, M., and Sherry, J. (2012). Haemocytic leukemia in Prince Edward Island (PEI) soft shell clam (*Mya arenaria*): spatial distribution in agriculturally impacted estuaries. *Sci. Total Environ.* 424, 130–142.
- Oprandy, J.J., and Chang, P.W. (1983). 5-bromodeoxyuridine induction of hematopoietic neoplasia and retrovirus activation in the soft-shell clam, *Mya arenaria*. *J. Invertebr. Pathol.* 42, 196–206.
- Oprandy, J.J., Chang, P.W., Pronovost, A.D., Cooper, K.R., Brown, R.S., and Yates, V.J. (1981). Isolation of a viral agent causing hematopoietic neoplasia in the soft-shell clam, *Mya arenaria*. *J. Invertebr. Pathol.* 38, 45–51.
- Pearse, A.M., and Swift, K. (2006). Allograft theory: transmission of devil facial-tumour disease. *Nature* 439, 549.
- Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Reno, P.W., House, M., and Illingworth, A. (1994). Flow cytometric and chromosome analysis of softshell clams, *Mya arenaria*, with disseminated neoplasia. *J. Invertebr. Pathol.* 64, 163–172.
- Rodzen, J.A., and May, B. (2002). Inheritance of microsatellite loci in the white sturgeon (*Acipenser transmontanus*). *Genome* 45, 1064–1076.
- Siah, A., McKenna, P., Danger, J.M., Johnson, G.R., and Berthe, F.C. (2011). Induction of transposase and polyprotein RNA levels in disseminated neoplastic hemocytes of soft-shell clams: *Mya arenaria*. *Dev. Comp. Immunol.* 35, 151–154.
- Siddle, H.V., and Kaufman, J. (2015). Immunology of naturally transmissible tumours. *Immunology* 144, 11–20.
- Siddle, H.V., Kreiss, A., Tovar, C., Yuen, C.K., Cheng, Y., Belov, K., Swift, K., Pearse, A.M., Hamede, R., Jones, M.E., et al. (2013). Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. *Proc. Natl. Acad. Sci. USA* 110, 5103–5108.
- Smolowitz, R.M., Miosky, D., and Reinisch, C.L. (1989). Ontogeny of leukemic cells of the soft shell clam. *J. Invertebr. Pathol.* 53, 41–51.
- St-Onge, P., Sevigny, J.M., Strasser, C., and Tremblay, R. (2013). Strong population differentiation of softshell clams (*Mya arenaria*) sampled across seven biogeographic marine ecoregions: possible selection and isolation by distance. *Mar. Biol.* 160, 1065–1081.

- Strasser, C.A., and Barber, P.H. (2009). Limited genetic variation and structure in softshell clams (*Mya arenaria*) across their native and introduced range. *Conserv. Genet.* **10**, 803–814.
- Sunila, I., and Farley, C.A. (1989). Environmental Limits for Survival of Sarcoma-Cells from the Soft-Shell Clam *Mya-Arenaria*. *Dis. Aquat. Organ.* **7**, 111–115.
- Taraska, N.G., and Anne Böttger, S. (2013). Selective initiation and transmission of disseminated neoplasia in the soft shell clam *Mya arenaria* dependent on natural disease prevalence and animal size. *J. Invertebr. Pathol.* **112**, 94–101.
- Vassilenko, E., and Baldwin, S.A. (2014). Using flow cytometry to detect haemic neoplasia in mussels (*Mytilus trossulus*) from the Pacific Coast of Southern British Columbia, Canada. *J. Invertebr. Pathol.* **117**, 68–72.
- Vassilenko, E.I., Muttray, A.F., Schulte, P.M., and Baldwin, S.A. (2010). Variations in p53-like cDNA sequence are correlated with mussel haemic neoplasia: A potential molecular-level tool for biomonitoring. *Mutat. Res.* **701**, 145–152.
- Voskoboinik, A., Newman, A.M., Corey, D.M., Sahoo, D., Pushkarev, D., Neff, N.F., Passarelli, B., Koh, W., Ishizuka, K.J., Palmeri, K.J., et al. (2013). Identification of a colonial chordate histocompatibility gene. *Science* **341**, 384–387.
- Walker, C., Böttger, S., and Low, B. (2006). Mortalin-based cytoplasmic sequestration of p53 in a nonmammalian cancer model. *Am. J. Pathol.* **168**, 1526–1530.
- Walker, C., Böttger, S.A., Mulkern, J., Jerszyk, E., Litvaitis, M., and Lesser, M. (2009). Mass culture and characterization of tumor cells from a naturally occurring invertebrate cancer model: applications for human and animal disease and environmental health. *Biol. Bull.* **216**, 23–39.
- Weinberg, J.R., Leavitt, D.F., Lancaster, B.A., and Capuzzo, J.M. (1997). Experimental field studies with *Mya arenaria* (Bivalvia) on the induction and effect of hematopoietic neoplasia. *J. Invertebr. Pathol.* **69**, 183–194.
- Yang, T.J., Chandler, J.P., and Dunne-Anway, S. (1987). Growth stage dependent expression of MHC antigens on the canine transmissible venereal sarcoma. *Br. J. Cancer* **55**, 131–134.
- Yevich, P.P., and Barszcz, C.A. (1978). Neoplasia in soft-shell clams (*Mya arenaria*) collected from oil-impacted sites. *Ann. N Y Acad. Sci.* **298**, 409–426.

# Indigenous Bacteria from the Gut Microbiota Regulate Host Serotonin Biosynthesis

## Graphical Abstract



## Authors

Jessica M. Yano, Kristie Yu, ..., Sarkis K. Mazmanian, Elaine Y. Hsiao

## Correspondence

ehsiao@caltech.edu

## In Brief

Indigenous spore-forming microbes from the gut microbiota produce metabolites that promote host serotonin biosynthesis in the gastrointestinal tract and impact gastrointestinal motility and hemostasis.

## Highlights

- Gut microbes regulate levels of 5-HT in the colon and blood
- Spore-forming bacteria modulate metabolites that promote colon 5-HT biosynthesis
- Microbiota-dependent changes in 5-HT impact GI motility and hemostasis
- Altering the microbiota could improve 5-HT-related disease symptoms





# Indigenous Bacteria from the Gut Microbiota Regulate Host Serotonin Biosynthesis

Jessica M. Yano,<sup>1</sup> Kristie Yu,<sup>1</sup> Gregory P. Donaldson,<sup>1</sup> Gauri G. Shastri,<sup>1</sup> Phoebe Ann,<sup>1</sup> Liang Ma,<sup>2</sup> Cathryn R. Nagler,<sup>3</sup> Rustem F. Ismagilov,<sup>2</sup> Sarkis K. Mazmanian,<sup>1</sup> and Elaine Y. Hsiao<sup>1,\*</sup>

<sup>1</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

<sup>2</sup>Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

<sup>3</sup>Department of Pathology and Department of Medicine, University of Chicago, Chicago, IL 60637, USA

\*Correspondence: [ehsiao@caltech.edu](mailto:ehsiao@caltech.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.047>

## SUMMARY

The gastrointestinal (GI) tract contains much of the body's serotonin (5-hydroxytryptamine, 5-HT), but mechanisms controlling the metabolism of gut-derived 5-HT remain unclear. Here, we demonstrate that the microbiota plays a critical role in regulating host 5-HT. Indigenous spore-forming bacteria (Sp) from the mouse and human microbiota promote 5-HT biosynthesis from colonic enterochromaffin cells (ECs), which supply 5-HT to the mucosa, lumen, and circulating platelets. Importantly, microbiota-dependent effects on gut 5-HT significantly impact host physiology, modulating GI motility and platelet function. We identify select fecal metabolites that are increased by Sp and that elevate 5-HT in chromaffin cell cultures, suggesting direct metabolic signaling of gut microbes to ECs. Furthermore, elevating luminal concentrations of particular microbial metabolites increases colonic and blood 5-HT in germ-free mice. Altogether, these findings demonstrate that Sp are important modulators of host 5-HT and further highlight a key role for host-microbiota interactions in regulating fundamental 5-HT-related biological processes.

## INTRODUCTION

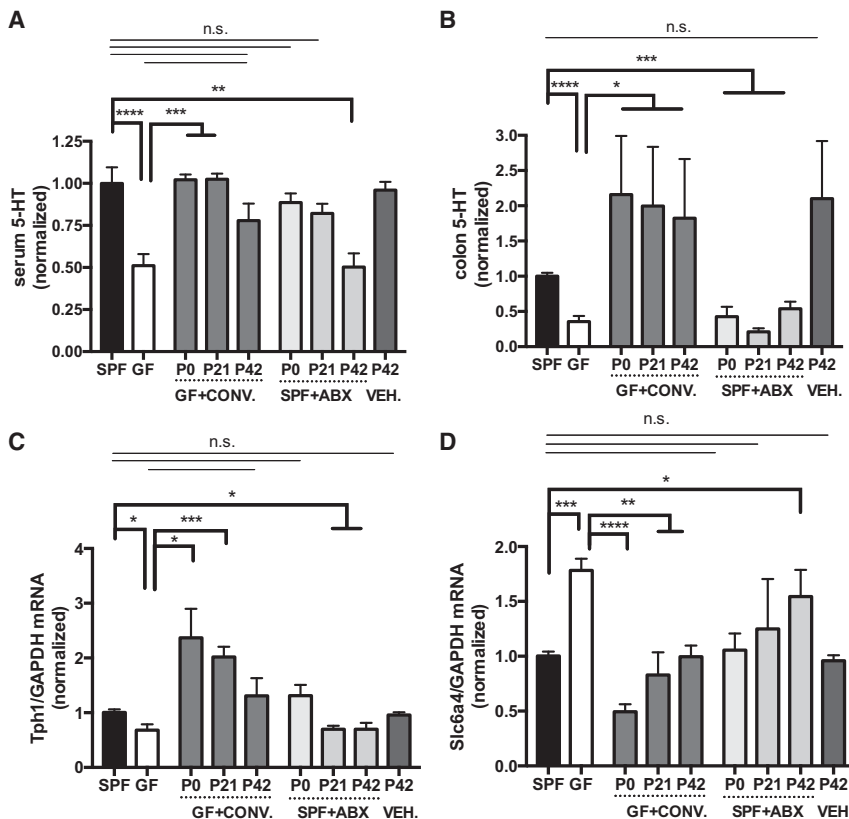
In addition to its role as a brain neurotransmitter, the monoamine serotonin (5-hydroxytryptamine [5-HT]) is an important regulatory factor in the gastrointestinal (GI) tract and other organ systems. More than 90% of the body's 5-HT is synthesized in the gut, where 5-HT activates as many as 14 different 5-HT receptor subtypes (Gershon and Tack, 2007) located on enterocytes (Hoffman et al., 2012), enteric neurons (Mawe and Hoffman, 2013), and immune cells (Baganz and Blakely, 2013). In addition, circulating platelets sequester 5-HT from the GI tract, releasing it to promote hemostasis and distributing it to various body sites (Amireault et al., 2013). As such, gut-derived 5-HT regulates diverse functions, including enteric motor and secretory reflexes (Gershon and Tack, 2007), platelet aggregation (Mercado et al.,

2013), immune responses (Baganz and Blakely, 2013), and bone development (Chabbi-Achengli et al., 2012; Yadav et al., 2008), and cardiac function (Côté et al., 2003). Furthermore, dysregulation of peripheral 5-HT is implicated in the pathogenesis of several diseases, including irritable bowel syndrome (IBS) (Stasi et al., 2014), cardiovascular disease (Ramage and Villalón, 2008), and osteoporosis (Ducy and Karsenty, 2010).

The molecular mechanisms controlling the metabolism of gut 5-HT remain unclear. In the GI tract, 5-HT is synthesized by specialized endocrine cells, called enterochromaffin cells (ECs), as well as mucosal mast cells and myenteric neurons (Gershon and Tack, 2007), but the functions of these different pools of gut 5-HT are incompletely understood. In addition, two different isoenzymes of tryptophan hydroxylase (Tph), Tph1 and Tph2, mediate non-neuronal versus neuronal 5-HT biosynthesis (Walther et al., 2003), but little is known regarding the endogenous signals that regulate Tph expression and activity.

Mammals are colonized by a vast and diverse collection of microbes that critically influences health and disease. Recent studies highlight a role for the microbiota in regulating blood 5-HT levels, wherein serum concentrations of 5-HT are substantially reduced in mice reared in the absence of microbial colonization (germ-free [GF]), compared to conventionally-colonized (specific pathogen-free [SPF]) controls (Sjögren et al., 2012; Wikoff et al., 2009). In addition, intestinal ECs are morphologically larger in GF versus SPF rats (Uribe et al., 1994), which suggests that microbes could impact the development and/or function of 5-HT-producing cells. Interestingly, some species of bacteria grown in culture can produce 5-HT (Tsavkelova et al., 2006), raising the question of whether indigenous members of the microbiota contribute to host 5-HT levels through de novo synthesis. Based on this emerging link between the microbiota and serum 5-HT concentrations, we aimed to determine how pathways of 5-HT metabolism are affected by the gut microbiota, to identify specific microbial communities and factors involved in conferring serotonergic effects, and to evaluate how microbial modulation of peripheral 5-HT impacts host physiology.

Here, we show that the microbiota promotes 5-HT biosynthesis from colonic ECs in a postnatally inducible and reversible manner. Spore-forming microbes (Sp) from the healthy mouse and human microbiota sufficiently mediate microbial effects on serum, colon, and fecal 5-HT levels. We further explore potential host-microbial interactions that regulate peripheral 5-HT by



**Figure 1. The Gut Microbiota Modulates Host Peripheral Serotonin Levels**

(A) Levels of serum 5-HT. Data are normalized to serum 5-HT in SPF mice ( $n = 8-13$ ).

(B) Levels of colon 5-HT relative to total protein. Data are normalized to colon 5-HT relative to total protein in SPF mice ( $n = 8-13$ ).

(C) Colonic expression of *TPH1* relative to *GAPDH*. Data are normalized to expression levels in SPF mice ( $n = 4$ ).

(D) Colonic expression of *SLC6A4* relative to *GAPDH*. Data are normalized to expression levels in SPF mice ( $n = 4$ ).

Data are presented as mean  $\pm$  SEM. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . n.s., not statistically significant; SPF, specific pathogen-free (conventionally-colonized); GF, germ-free; CONV., SPF conventionalized; ABX, antibiotic-treated; VEH, vehicle (water)-treated. See also Figure S1.

surveying microbial influences on the fecal metabolome. We find that particular microbial metabolites are elevated by Sp and likely signal directly to colonic ECs to promote 5-HT biosynthesis. Importantly, microbiota-mediated changes in colonic 5-HT regulate GI motility and hemostasis in the host, suggesting that targeting the microbiota can serve as a tractable approach for modulating peripheral 5-HT bioavailability and treating 5-HT-related disease symptoms.

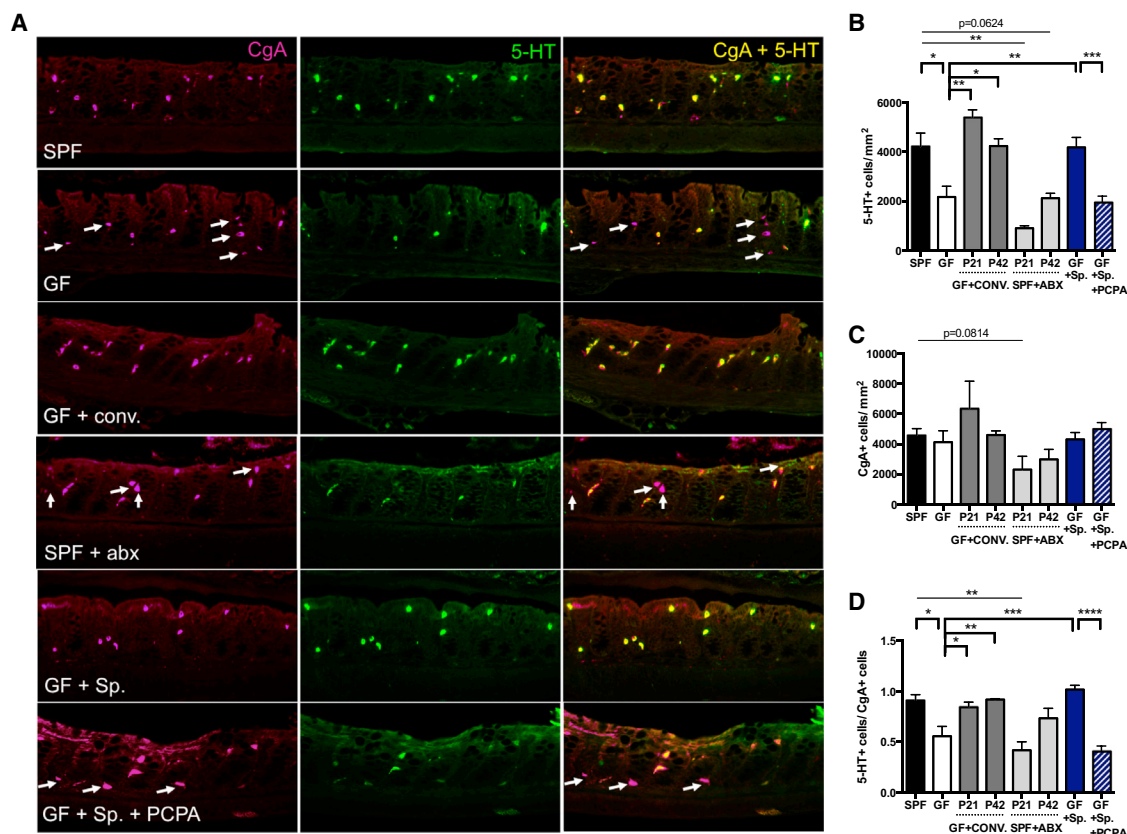
## RESULTS

### The Gut Microbiota Modulates Host Peripheral Serotonin Levels

Adult GF mice display deficient serum (Sjögren et al., 2012; Wikoff et al., 2009) (Figure 1A) and plasma (Figure S1A) 5-HT concentrations compared to SPF controls, but the cellular sources of this disruption are undefined. Consistent with the understanding that much of the body's 5-HT derives from the GI tract, we find that GF mice exhibit significantly decreased levels of colonic and fecal 5-HT compared to SPF controls (Figures 1B and S1A; Table S1). This deficit in 5-HT is observed broadly across the distal, medial and proximal colon (Figure S1D), but not in the small intestine (Figures S1A, S2A, and S2B), suggesting a specific role for the microbiota in regulating colonic 5-HT. Decreased levels of 5-HT are localized to colonic chromogranin A-positive (CgA+) enterochromaffin cells (ECs) (Figure 2), and not to small intestinal ECs (Figures S2A and S2B). Low 5-HT signal is seen in both GF and SPF colonic mast cells and enteric

neurons (Figure 2A), which are minor producers of 5-HT (Gershon and Tack, 2007). There is no difference between adult GF and SPF mice in the abundance of CgA+ ECs (Figure 2C), suggesting that decreases in colon 5-HT result from abnormal 5-HT metabolism rather than impaired development of ECs.

To identify the specific steps of 5-HT metabolism that are affected by the microbiota, key intermediates of the 5-HT pathway were assessed in colons from GF versus SPF mice. We find that GF colons exhibit decreased expression of *TPH1* (Figures 1C and S1D) (Sjögren et al., 2012), the rate-limiting enzyme for 5-HT biosynthesis in ECs, but no difference in expression of enzymes involved in 5-HT packaging, release and catabolism (Figure S1C). GF mice also display elevated colonic expression of the 5-HT transporter *SLC6A4* (Figures 1D and S1E) (Sjögren et al., 2012), synthesized broadly by enterocytes to enable 5-HT uptake (Wade et al., 1996). This could reflect a compensatory response to deficient 5-HT synthesis by host ECs, based on the finding that chemical Tph inhibition modulates *SLC6A4* expression (Figures S2C and S2D). There is no difference between GF and SPF mice in colonic expression of neural-specific isoforms of 5-HT enzymes (Figure S1F), consistent with data showing no apparent difference in 5-HT-specific staining in enteric neurons (Figure 2). Despite deficient levels of colon, fecal, and serum 5-HT (Figures 1A, 1B, and S1A; Table S1), GF mice exhibit significantly increased levels of the Tph substrate, tryptophan (Trp), in both feces (Table S1) and serum (Sjögren et al., 2012; Wikoff et al., 2009), suggesting that primary disruptions in host *TPH1* expression result in Trp accumulation. Oral supplementation of GF mice with the Tph product, 5-hydroxytryptophan (5-HTP), sufficiently ameliorates deficits in colon and serum 5-HT, whereas supplementation with the Tph substrate Trp has no restorative effect (Figures S1G–S1I). Collectively, these data support the notion that the microbiota promotes 5-HT biosynthesis by elevating *TPH1* expression in colonic ECs.



**Figure 2. Indigenous Spore-Forming Bacteria Increase 5-HT Levels in Colon Enterochromaffin Cells**

(A) Representative images of colons stained for chromogranin A (CgA) (left), 5-HT (center), and merged (right). Arrows indicate CgA-positive cells that lack 5-HT staining ( $n = 3-7$  mice/group).

(B) Quantitation of 5-HT+ cell number per area of colonic epithelial tissue ( $n = 3-7$  mice/group).

(C) Quantitation of CgA+ cell number per area of colonic epithelial tissue ( $n = 3-7$  mice/group).

(D) Ratio of 5-HT+ cells/CgA+ cells per area of colonic epithelial tissue ( $n = 3-7$  mice/group).

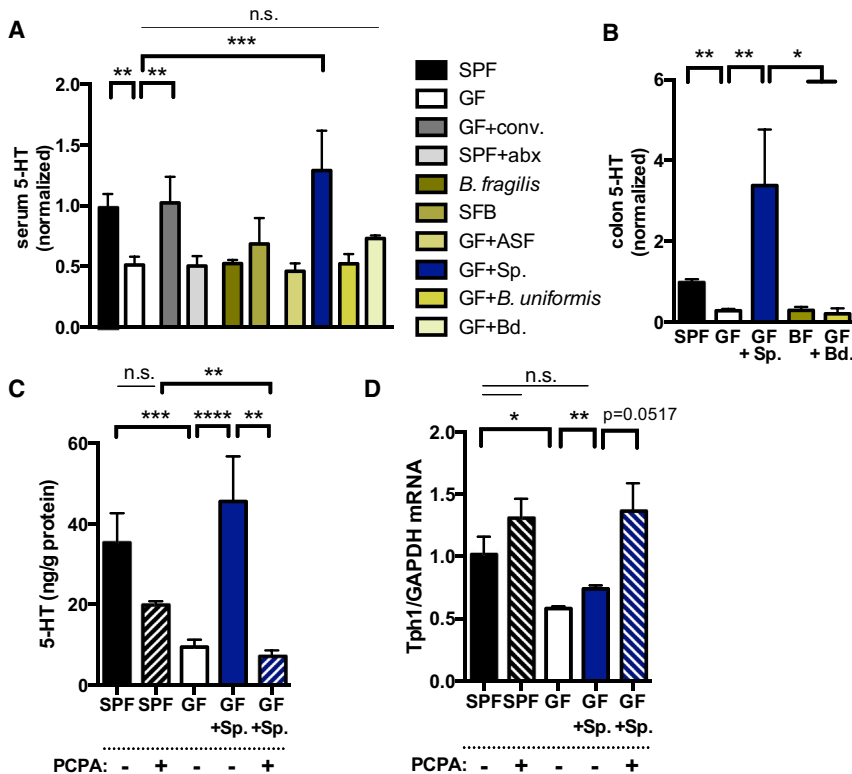
Data are presented as mean  $\pm$  SEM. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . SPF, specific pathogen-free (conventionally-colonized); GF, germ-free; CONV., SPF conventionalized; ABX, antibiotic-treated; Sp., spore-forming bacteria; PCPA, para-chlorophenylalanine.

See also Figure S2.

To confirm that deficient 5-HT levels in GF mice are microbiota-dependent and further determine whether effects are age-dependent, GF mice were conventionalized with an SPF microbiota at birth (postnatal day [P] 0), weaning (P21), or early adulthood (P42) and then evaluated at P56 for levels of 5-HT and expression of 5-HT-related genes. GF mice conventionalized at each age with an SPF microbiota exhibit restored serum (Figure 1A) and colon (Figure 1B) 5-HT levels, with more pronounced effects seen at earlier ages of colonization. Colonic expression of *TPH1* and *SLC6A4* is similarly corrected by postnatal conventionalization of GF mice (Figures 1C and 1D), with more substantial changes from P0 conventionalization. Increases in 5-HT are localized to colonic ECs (Figure 2). These findings indicate that postnatal reconstitution of the gut microbiota can correct the 5-HT deficiency seen in GF mice and further suggest that gut microbes exert a continuous effect on 5-HT synthesis by modulating EC function. Overall, we demonstrate that microbiota-mediated elevation of host 5-HT is postnatally inducible, persistent from the time of

conventionalization and not dependent on the timing of host development.

To assess the reversibility of microbial effects on host 5-HT metabolism, we depleted the gut microbiota in SPF mice via bi-daily antibiotic treatment beginning on P0, P21, or P42 and until P56. Treatment of P42 SPF mice with a cocktail of ampicillin, vancomycin, neomycin, and metronidazole (Reikvam et al., 2011) sufficiently recapitulates GF-associated deficits in serum and colon 5-HT and alterations in host colonic *TPH1* and *SLC6A4* expression (Figures 1 and 2). Interestingly, P0 and P21 antibiotic treatment also induces GF-related deficits in colonic 5-HT, but the effects on serum 5-HT are more pronounced when administered at P42, compared to P0 and P21 (Figure 1), suggesting potential confounding effects of early life or prolonged antibiotic treatment on microbiota-mediated modulation of peripheral 5-HT. Antibiotics can elicit several direct effects on host cells (Shimizu et al., 2003; Westphal et al., 1994), which may underlie differences between P0 treatment and GF status. That P42 antibiotic treatment of SPF mice results in



**Figure 3. Indigenous Spore-Forming Bacteria Induce Colon 5-HT Biosynthesis and Systemic 5-HT Bioavailability**

(A) Levels of serum 5-HT. Data are normalized to serum 5-HT levels in SPF mice. SPF,  $n = 13$ ; GF,  $n = 17$ ; GF+conv, P21 conventionalization,  $n = 4$ ; SPF+Abx, P42 antibiotic treatment,  $n = 7$ ; *B. fragilis* monoassociation (BF),  $n = 6$ ; SFB, Segmented Filamentous Bacteria monoassociation,  $n = 4$ ; ASF, Altered Schaedler Flora P21 colonization,  $n = 4$ ; Sp, spore-forming bacteria, P21 colonization,  $n = 4$ ; *B. uniformis* P21 colonization,  $n = 4$ ; Bd, *Bacteroides* consortium,  $n = 3$ .

(B) Levels of colon 5-HT relative to total protein. Data are normalized to colon 5-HT relative to total protein in SPF mice ( $n = 5-15$ ).

(C) Levels of colon 5-HT relative to total protein after intrarectal treatment with the Tph inhibitor, PCPA, or vehicle ( $n = 4$ ).

(D) Colonic expression of *TPH1* relative to *GAPDH*. Data are normalized to expression levels in SPF mice ( $n = 3$ ).

Data are presented as mean  $\pm$  SEM. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ . n.s., not statistically significant; SPF, specific pathogen-free (conventionally-colonized); GF, germ-free; Sp, spore-forming bacteria; PCPA, para-chlorophenylalanine.

See also Figure S3.

5-HT phenotypes analogous to those seen in GF mice demonstrates that microbiota effects on host 5-HT can be abrogated postnatally and further supports the plasticity of 5-HT modulation by indigenous gut microbes. Altogether, these data indicate that the gut microbiota plays a key role in raising levels of colon and serum 5-HT, by promoting 5-HT in colonic ECs in an inducible and reversible manner.

### Indigenous Spore-Forming Microbes Promote Host Serotonin Biosynthesis

In light of our finding that 5-HT levels are decreased in colons but not small intestines of GF mice compared to SPF controls, we hypothesized that specific subsets of gut microbes are responsible for affecting host 5-HT pathways. Mice mono-colonized with *Bacteroides fragilis* or Segmented Filamentous Bacteria (SFB) display deficits in serum 5-HT that are comparable to those seen in GF mice (Figure 3A). Moreover, postnatal colonization (P42) with *Bacteroides uniformis*, altered Schaedler flora (ASF), an eight-microbe consortium known to correct gross intestinal pathology in GF mice (Dewhirst et al., 1999), or with cultured *Bacteroides spp.* from the SPF mouse microbiota, has no significant effect on the 5-HT deficiency seen in GF mice (Figures 3A and 3B). Interestingly, however, GF mice colonized at P42 with indigenous spore-forming microbes from the mouse SPF microbiota (Sp), known to be dominated by Clostridial species (Atarashi et al., 2013; Stefk et al., 2014) (Table S2), exhibit complete restoration of serum and colon 5-HT to levels observed in SPF mice (Figures 3A and 3B). Consistent with this, Sp colonization of GF mice increases 5-HT staining

colocalized to CgA+ ECs (Figure 2), elevates host colonic *TPH1* expression (Figure 3D) and decreases *SLC6A4* expression (Figure 3E) toward levels seen in SPF mice. Improvements in serum 5-HT are observed within 2 days after inoculation of GF mice with Sp (Figure S2E) and do not correlate with amelioration of abnormal cecal weight (Figure S2F). Importantly, Sp also elevates colonic 5-HT in Rag1 knockout mice (Figure S2G), which lack adaptive immune cells, indicating that the effects of Sp on gut 5-HT are not dependent on Sp-mediated regulatory T cell induction (Stefk et al., 2014). Notably, the 5-HT-promoting effects of Sp are recapitulated by colonization of GF mice with spore-forming microbes from the healthy human colonic microbiota (hSp) (Figure S3), suggesting that the serotonergic function of this community is conserved across mice and humans.

To determine whether the effects of Sp on host 5-HT depend on colonic Tph activity, we colonized GF mice with Sp on P42 and then administered the Tph inhibitor para-chlorophenylalanine (PCPA) intrarectally twice daily for 3 days prior to 5-HT assessments on P56 (Liu et al., 2008). Intrarectal injection of PCPA sufficiently blocks the ability of Sp to elevate colon and serum 5-HT levels (Figures 3C and S2C), as well as Sp-mediated increases in 5-HT staining in ECs (Figure 2). Similar effects of PCPA treatment on blocking increases in colon 5-HT, serum 5-HT, and 5-HT staining in colonic ECs are seen in GF mice colonized with hSp (Figure S3). Interestingly, inhibiting Tph activity with PCPA results in a compensatory increase in colonic *TPH1* and decrease in *SLC6A4* (Figures 3D and S2D) expression in Sp-colonized mice, supporting the notion that



microbiota-dependent changes in 5-HT transporter levels occur as a secondary response to Tph modulation.

To further evaluate whether changes in *SLC6A4* expression are necessary for microbiota-mediated alterations in peripheral 5-HT, we tested the effects of microbiota manipulations on colon and serum 5-HT in *SLC6A4* heterozygous (+/−) and complete (−/−) knockout (KO) mice. Depleting the microbiota via P42-P56 antibiotic treatment (Reikvam et al., 2011) of SPF *SLC6A4*<sup>+/−</sup> and *SLC6A4*<sup>−/−</sup> mice effectively decreases colonic 5-HT levels (Figures S4A and S4B), indicating that the microbiota is required for promoting gut 5-HT in *Slc6a4*-deficient mice. Colonizing antibiotic-treated *SLC6A4*<sup>+/−</sup> and *SLC6A4*<sup>−/−</sup> mice with Sp raises colon 5-HT to levels seen in untreated SPF *SLC6A4*<sup>+/−</sup> and *SLC6A4*<sup>−/−</sup> mice (Figure S4A), demonstrating that *Slc6a4* is not required for conferring the effects of Sp on gut 5-HT. Antibiotic-induced decreases and Sp-induced increases in colon 5-HT levels can be attributed to modulation of 5-HT content in colonic ECs from *SLC6A4*<sup>+/−</sup> and *SLC6A4*<sup>−/−</sup> mice (Figure S4C). Similar effects of antibiotic treatment and Sp colonization are seen for serum 5-HT in *SLC6A4*<sup>+/−</sup> mice, whereas *SLC6A4*<sup>−/−</sup> mice exhibit low to undetectable levels of serum 5-HT, highlighting the dependence of platelets on *Slc6a4*-mediated 5-HT uptake (Figure S4B). Taken together, these data support a role for Sp in promoting Tph1-mediated 5-HT biosynthesis by colonic ECs, regulating both colon and serum levels of 5-HT.

### Microbiota-Mediated Regulation of Host Serotonin Modulates Gastrointestinal Motility

Intestinal 5-HT plays an important role in stimulating the enteric nervous system and GI function (Gershon and Tack, 2007). To determine whether microbiota-dependent modulation of colonic 5-HT impacts GI motility, we colonized P42 GF mice with Sp and then tested for GI transit and colonic neuronal activation at P56. Sp colonization ameliorates GF-associated abnormalities in GI motility, significantly decreasing total transit time and increasing the rate of fecal output in a Tph-dependent manner (Figures 4A and 4B). Similar effects are seen in *SLC6A4*<sup>+/−</sup> and *SLC6A4*<sup>−/−</sup> mice, where Sp colonization of antibiotic-treated mice restores GI transit time toward levels seen in untreated SPF *SLC6A4*<sup>+/−</sup> and *SLC6A4*<sup>−/−</sup> controls (Figure S4E).

Consistent with deficits in GI motility, steady-state activation of 5-HT receptor subtype 4 (5HT4)-expressing cells in the colonic submucosa and muscularis externa is decreased in GF mice compared to SPF controls, as measured by colocalized expression of 5HT4 with the immediate early gene, *c-fos* (Figures 4C–4E). Colonization of GF mice with Sp increases 5HT4+ *c-fos*+ staining to levels seen in SPF mice, and this effect is dependent on colonic Tph activity (Figures 4C–4E), which aligns well with the understanding that Sp-induced elevations in colonic 5-HT promote GI motility by activation of 5HT4+ enteric neurons (Mawe and Hoffman, 2013). In addition, colonic activation of intrinsic afferent primary neurons (IPANs) of the myenteric plexus is decreased in GF mice (McVey Neufeld et al., 2013) and improved by colonization with Sp, as measured by colocalization of *c-fos* and the IPAN marker, calcitonin (Calb2) (Figure 4F). Inhibiting Tph activity with PCPA decreases IPAN activation in Sp-colonized mice, suggesting that some IPAN responses to Sp depend on host 5-HT synthesis (Figure 4F). Altogether, these

findings indicate that Sp-mediated increases in colonic 5-HT biosynthesis are important for gut sensorimotor function.

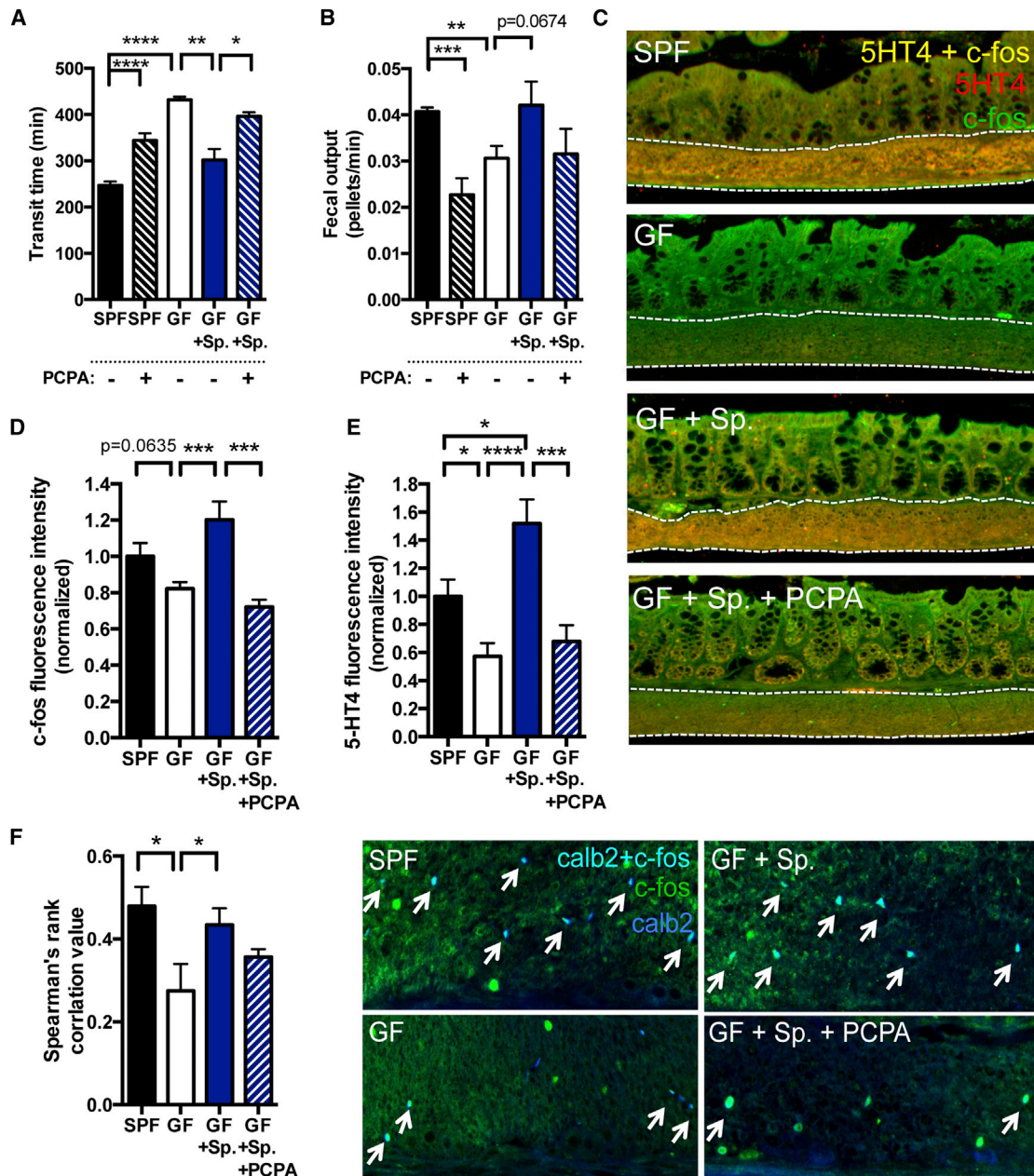
### Microbiota-Mediated Regulation of Host Serotonin Modulates Platelet Function

Platelets uptake gut-derived 5-HT and release it at sites of vessel injury to promote blood coagulation. To determine if microbiota-dependent modulation of colon (Figures 1 and 3) and plasma (Figure S1A) 5-HT impacts platelet function, we colonized P42 mice with Sp and then examined blood clotting, platelet activation and platelet aggregation at P56. In a tail bleed assay (Liu et al., 2012), GF mice exhibit trending increases in time to cessation of bleeding compared to SPF mice, suggesting impaired blood coagulation (Figure 5A). Colonization of GF mice with Sp ameliorates abnormalities in bleeding time to levels seen in SPF controls, and this effect is attenuated by intrarectal administration of PCPA (Figure 5A), indicating that Sp-mediated improvements in coagulation may be dependent on colonic Tph activity. Notably, the impact of acute colonic PCPA treatment on reducing 5-HT content and 5-HT-related functions in platelets may be tempered by the fact that mouse platelets have a lifespan of ~4 days (Odell and McDonald, 1961). There were no significant differences between treatment groups in total platelet counts (Figure S5A).

In light of inherent limitations of the tail bleed assay (Liu et al., 2012), we focused subsequent experiments particularly on platelet activity. Platelets isolated from GF mice display decreased activation in response to in vitro type I fibrillar collagen stimulation, as measured by reduced surface expression of the activation markers granulophysin (CD63), P-selectin, and JON/A (integrin  $\alpha$ IIb $\beta$ 3) (Figures 5D–5F) (Ziu et al., 2012). Sp colonization of GF mice leads to partial restoration in the expression of platelet activation markers, and this effect depends on colonic Tph activity (Figures 5D–5F). Moreover, platelets isolated from GF mice exhibit impaired aggregation in response to in vitro collagen stimulation, as measured by decreased levels of high granularity, high mass aggregates detected by both flow cytometry (De Cuyper et al., 2013; Nieswandt et al., 2004) (Figures 5B, 5C, S5C, and S5D) and imaging (Figure S5B). Colonization of GF mice with Sp restores levels of platelet aggregation to those seen in SPF mice. These effects of Sp on correcting impaired platelet aggregation are attenuated by colonic PCPA injection, indicating dependence on Tph activity. Overall, these findings suggest that Sp-mediated elevations in colonic 5-HT, and thus platelet 5-HT, promote platelet activation and aggregation relevant to hemostasis.

### Microbial Metabolites Mediate Effects of the Microbiota on Host Serotonin

In light of the important role for Sp in regulating 5-HT-related intestinal and platelet function, we aimed to identify specific microbial factors responsible for conferring the serotonergic effects of Sp. Based on our finding that Sp elevates 5-HT particularly in colonic ECs (Figure 2), we hypothesized that Sp promotes levels of a soluble factor that signals directly to ECs to modulate *TPH1* expression and 5-HT biosynthesis. To test this, we prepared filtrates of total colonic luminal contents from Sp-colonized mice and controls and evaluated their effects on levels of 5-HT in RIN14B chromaffin cell cultures (Nozawa et al., 2009). Relative



**Figure 4. Microbiota-Mediated Regulation of Host Serotonin Modulates Gastrointestinal Motility**

(A) Total time for transit of orally administered carmine red solution through the GI tract (n = 4–8).

(B) Defecation rate as measured by number of fecal pellets produced relative to total transit time (n = 4–8).

(C) Representative images of c-fos and 5HT4 colocalization in the colonic submucosa and muscularis externa (n = 4–5 mice/group).

(D) Quantitation of total c-fos fluorescence intensity in the colonic submucosa and muscularis externa (n = 4–5 mice/group).

(E) Quantitation of total 5HT4 fluorescence intensity in the colonic submucosa and muscularis externa (n = 4–5 mice/group).

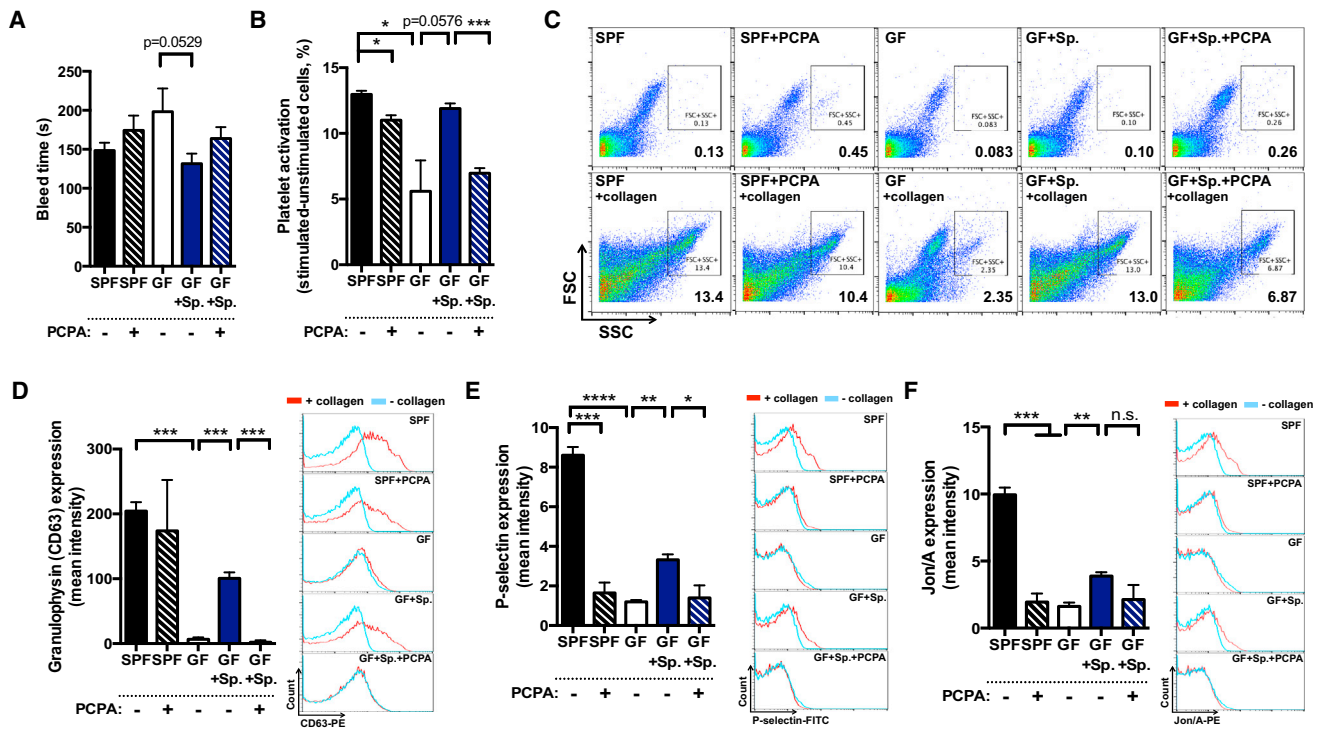
(F) Quantitation and representative images of c-fos and calb2 (calretinin) colocalization in the colonic submucosa and muscularis externa (n = 5–8 mice/group).

Data are presented as mean  $\pm$  SEM. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001. SPF, specific pathogen-free (conventionally-colonized); GF, germ-free; Sp, spore-forming bacteria; PCPA, para-chlorophenylalanine.

See also Figure S4.

to vehicle-treated controls, there is no significant effect of filtered colonic luminal contents from GF mice on levels of 5-HT released or *TPH1* expressed from RIN14B cells (Figures 6A and 6B).

Filtered colonic luminal contents from SPF and Sp-colonized mice sufficiently induce 5-HT from RIN14B cells (Figure 6A), to levels comparable to those elicited by the calcium ionophore,



**Figure 5. Microbiota-Mediated Regulation of Host Serotonin Modulates Hemostasis**

(A) Time to cessation of bleeding in response to tail injury (n = 7–16).

(B) Platelet activation, as measured by percentage of large, high granularity (FSC<sup>high</sup>, SSC<sup>high</sup>) events after collagen stimulation relative to unstimulated controls (n = 3).

(C) Representative flow cytometry plots of large, high granularity (FSC<sup>high</sup>, SSC<sup>high</sup>) activated platelets after collagen stimulation (bottom), as compared to unstimulated controls (top) (n = 3).

(D–F) Geometric mean fluorescence intensity of granulephysin (CD63) (D), P-selectin (E), and JON/A (integrin  $\alpha$ IIb $\beta$ 3) (F) expression in collagen-stimulated platelets (left). Representative histograms (right) of event count versus fluorescence intensity (log scale) for platelets treated with collagen (red line) or vehicle (blue line) (n = 3).

Data for platelet assays are representative of three independent trials with at least three mice in each group. Data are presented as mean  $\pm$  SEM. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001. n.s., not statistically significant; SPF, specific pathogen-free (conventionally-colonized); GF, germ-free; Sp, spore-forming bacteria; PCPA, para-chlorophenylalanine.

See also Figure S5.

ionomycin, as a positive control. *TPH1* expression is also elevated in chromaffin cells exposed to SPF and Sp luminal filtrates, suggesting increased 5-HT synthesis. This is in contrast to ionomycin, which stimulates 5-HT release, but has no effect on *TPH1* expression, from RIN14B cells. Importantly, these findings suggest that microbiota-mediated increases in gut 5-HT are conferred via direct signaling of a soluble, Sp-modulated factor to colonic ECs.

We utilized metabolomic profiling to identify candidate Sp-dependent, 5-HT-inducing molecules in feces from adult mice. Sp colonization of GF mice leads to statistically significant alterations in 75% of the 416 metabolites detected, of which 76% are elevated and 24% are reduced, relative to vehicle-treated GF controls (Tables S1 and S3). Similar changes are seen with hSp colonization, leading to co-clustering of Sp and hSp samples by principal components analysis (PCA) (Figure 6C). ASF colonization has a mild effect, significantly modulating 50% of metabolites detected (66% increased, 36% decreased) (Table S3), and forming a distinct but proximal cluster to GF controls

by PCA (Figure 6C). Postnatal conventionalization of GF mice with an SPF microbiota alters 66% of all metabolites detected (59% increased, 41% decreased) (Table S3) and produces substantial changes in the metabolome that are distinguishable from the effects of Sp, hSp, and ASF along PC2 (Figure 6C). Notably, Sp, hSp, and SPF colonization results in similar shifts along PC1, compared to vehicle and ASF-treated controls, suggesting common metabolic alterations among communities that similarly elevate peripheral 5-HT levels. Metabolomics profiling confirms that fecal 5-HT is commonly upregulated in the Sp, hSp, and SPF fecal metabolome and comparatively low in ASF and GF samples (Table S1). Simple linear regression reveals 83 metabolites that co-vary with 5-HT ( $r^2 \geq 0.25$ ), 47 of which correlate positively and 36 of which correlate negatively with 5-HT levels (Figure S6A; Table S4).

To determine whether specific metabolites mediate the effects of Sp on 5-HT, we tested a subset of biochemicals that were commonly upregulated by Sp, hSp, and SPF, and that positively correlated with 5-HT levels (Figure S6A; Table S4), for their ability



to induce 5-HT in vitro and in vivo. We also tested the short chain fatty acids, acetate, butyrate, and propionate, which were previously shown to be produced by Sp (Atarashi et al., 2013) and to stimulate 5-HT release from ECs (Fukumoto et al., 2003). Of 16 metabolites examined,  $\alpha$ -tocopherol, butyrate, cholate, deoxycholate, p-aminobenzoate (PABA), propionate, and tyramine elevate 5-HT in RIN14B chromaffin cell cultures (Figure 6D). Elevations in 5-HT correspond to increases in *TPH1* expression from RIN14B cells (Figure 6E), suggesting that particular metabolites induced by Sp enhance 5-HT biosynthesis by ECs. We further tested for sufficiency to induce 5-HT in vivo. Notably, raising luminal concentrations of deoxycholate in colons of GF mice to levels seen in SPF mice (Sayin et al., 2013) sufficiently increases colon and serum 5-HT compared to vehicle-injected controls (Figures 6F and S6B). This restoration of peripheral 5-HT correlates with elevations in colonic *TPH1* expression (Figure 6F). Increases in colon and serum 5-HT are also seen with injection of  $\alpha$ -tocopherol, PABA and tyramine into colons of GF mice (Figures S6B and S6C). Consistent with in vitro RIN14B data, oleanolate has no statistically significant effect on elevating colon or serum 5-HT in GF mice (Figures S6B and S6C). Importantly, the effects of a single rectal injection of deoxycholate or  $\alpha$ -tocopherol on raising colon 5-HT levels in GF mice are weak and transient, peaking within 1 hr of injection (Figure S6C). Consistent with this, there is no significant effect of acute colonic metabolite injection on GI transit time (Figure S6D), and there is only a trending improvement on platelet activation (Figure S6E). Our finding that Sp colonization leads to lasting increases in colon and blood 5-HT levels (Figure 3), and long-term changes in the fecal metabolome (Figure 6C; Tables S1 and S3), suggests that Sp colonization results in persistent elevations of 5-HT-modulating luminal metabolites. Future studies on whether chronic, colon-restricted increases in Sp-regulated metabolites sufficiently correct GI motility and platelet function in GF mice, and whether this occurs in a 5-HT-dependent manner, are warranted. In addition, we demonstrate that select concentrations of Sp-associated metabolites sufficiently promote 5-HT in vitro and in vivo, but whether the metabolites are necessary for mediating the serotonergic effects of Sp is unclear. Overall, these data reveal that indigenous spore-forming microbes promote 5-HT biosynthesis from colonic ECs, modulating 5-HT concentrations in both colon and blood. Furthermore, we identify select microbial metabolites that confer the serotonergic effects of indigenous spore-forming microbes, likely by signaling directly to colonic ECs to promote Tph1 expression and 5-HT biosynthesis.

## DISCUSSION

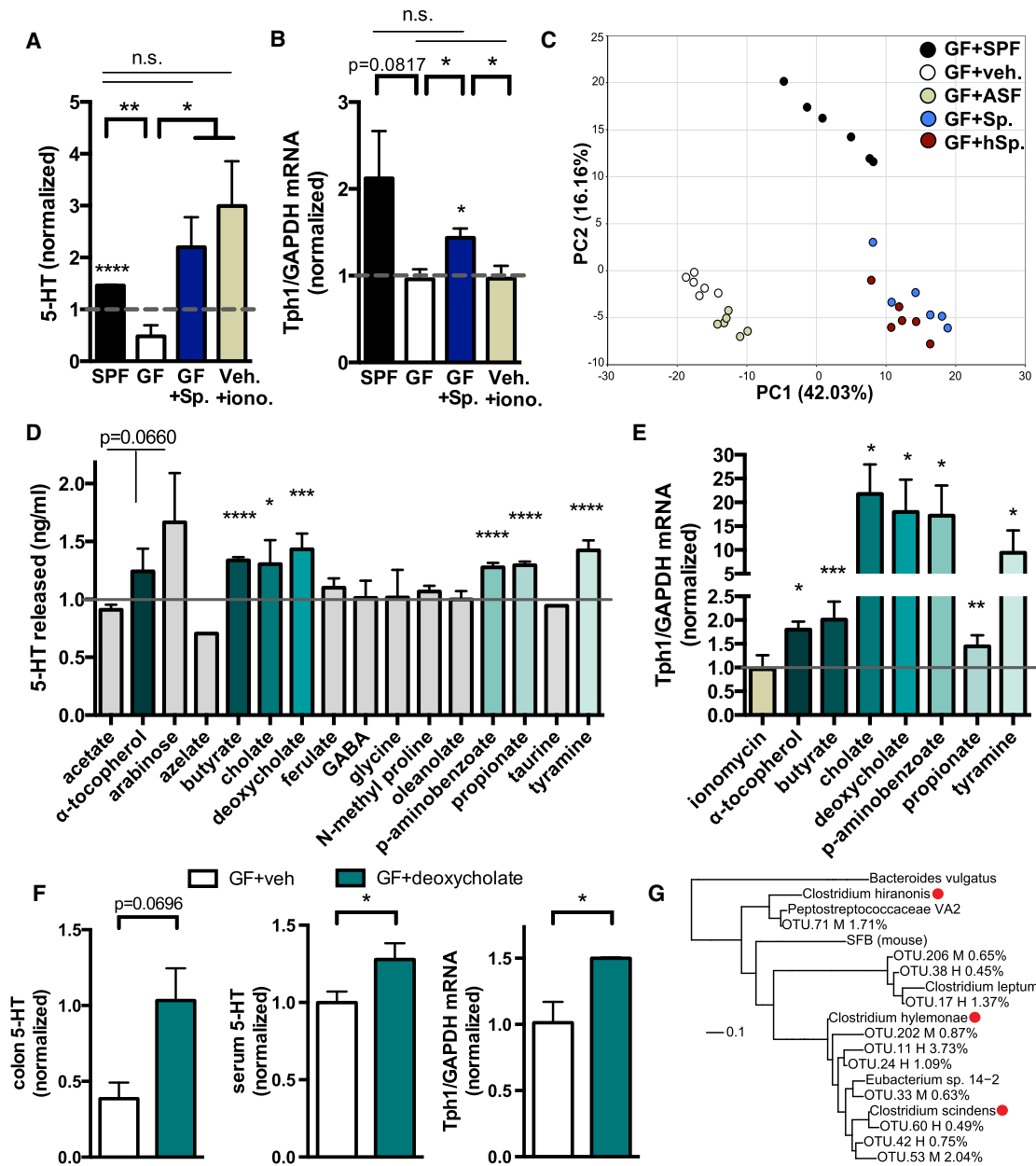
The GI tract is an important site for 5-HT biosynthesis, but the regulatory mechanisms underlying the metabolism of gut-derived 5-HT are incompletely understood. Here, we demonstrate that the gut microbiota plays a key role in promoting levels of colon and blood 5-HT, largely by elevating synthesis by host ECs. This host-microbiota interaction contributes to a growing appreciation that the microbiota regulates many aspects of GI physiology by signaling to host cells. Whether particular members of the microbiota contribute 5-HT by de novo synthesis remains unclear. Some bacteria, including *Corynebacterium* spp.,

*Streptococcus* spp., and *Escherichia coli*, are reported to synthesize 5-HT in culture (Roshchina, 2010), but this is believed to occur independently of Tph, by decarboxylation of tryptophan to tryptamine (Williams et al., 2014), as seen in plants (Oleskin et al., 1998). Our finding that colonic PCPA administration blocks the ability of the microbiota to promote colonic and blood 5-HT (Figures 3C and 3D) suggests that gut microbes require host Tph activity to upregulate peripheral 5-HT. Furthermore, SPF Tph1 KO mice lack >90% of intestinal and blood 5-HT levels (Savelieva et al., 2008), indicating that <10% of peripheral 5-HT is contributed directly by microbial synthesis or by Tph2-mediated biosynthesis in these mice. We find that the microbiota regulates relatively high levels of peripheral 5-HT, 64% of colonic (Figure 1), and 49% of serum concentrations (Figure 1) (Sjögren et al., 2012; Wikoff et al., 2009), further supporting the notion that the microbiota modulates 5-HT metabolism primarily by affecting host colonic ECs. Consistent with the understanding that ECs secrete low levels of 5-HT into the lumen, fecal concentrations of 5-HT are also significantly increased by the microbiota. Interestingly, 5-HT is reported to stimulate the growth of *Enterococcus faecalis*, *E. coli*, and *Rhodospirillum rubrum* in culture (Oleskin et al., 1998; Tsavkelova et al., 2006). In addition, 5-HT is a structural analog of auxins found in *E. faecalis*, *R. rubrum*, and *Staphylococcus aureus*, among other bacteria. Whether particular members of the microbiota alter host 5-HT biosynthesis to, in turn, support colonization, growth, or resilience of particular gut microbes is an interesting question for future study.

We demonstrate that indigenous spore-forming microbes from colons of SPF mice (Sp) and from a healthy human colon (hSp) sufficiently mediate microbiota effects on colonic and blood 5-HT. While we show that *B. fragilis*, *B. uniformis*, SFB, ASF, and a consortium of *Bacteroides* species cultured from mice, including *B. thetaiotaomicron*, *B. acidifaciens*, and *B. vulgatus*, have no effect on host peripheral 5-HT (Figure 3), whether other non-Sp microbial species or communities are capable of modulating colonic and serum 5-HT remains unclear. Interestingly, Sp and hSp are known to promote regulatory T cell levels in the colons, but not small intestines, of GF and SPF mice (Atarashi et al., 2013). This regional specificity is also seen with microbiota-induced 5-HT biosynthesis, which occurs in colonic, but not small intestinal, ECs (Figures S1A, S2A, and S2B). We find that Sp elevates colon 5-HT levels even in Rag1 KO mice (Figure S2G), indicating that the serotonergic effects of Sp are not dependent on T and B cells. Whether 5-HT modulation contributes to the immunosuppressive effects of Sp, however, is unclear. In light of increasing evidence that innate and adaptive immune cells express a variety of 5-HT receptors (Baganz and Blakely, 2013), future studies examining whether Sp-mediated increases in peripheral 5-HT levels impact cellular immune responses will be of interest.

Consistent with our finding that the microbiota modulates colon and serum 5-HT via interactions with host colonic ECs, we find that particular fecal metabolites are similarly elevated by SPF, Sp, and hSp microbiota and sufficiently promote 5-HT in chromaffin cell cultures and in vivo (Figure 6; Table S1). Deoxycholate is a secondary bile acid, produced by microbial biotransformation of cholate. In addition to facilitating lipid absorption, it has endocrine, immunological, and antibiotic effects and is





**Figure 6. Microbial Metabolites Mediate Effects of the Microbiota on Host Serotonin**

(A) Levels of 5-HT released from RIN14B cells after exposure to colonic luminal filtrate from SPF, GF, and Sp-colonized mice, or to ionomycin (iono). Data are normalized to 5-HT levels in vehicle-treated controls (hatched gray line at 1). Asterisks directly above bars indicate significance compared to controls; asterisks at the top of the graph denote significance between experimental groups (n = 3).

(B) Expression of *TPH1* relative to *GAPDH* in RIN14B cells after exposure to colon luminal filtrate from SPF, GF and Sp-colonized mice, or to ionomycin (iono). Data are normalized to gene expression in vehicle-treated controls (hatched gray line at 1). Asterisks directly above bars indicate significance compared to controls, whereas asterisks at the top of the graph denote significance between experimental groups (n = 4).

(C) Principal components analysis of the fecal metabolome from GF mice colonized with SPF, ASF, Sp, or hSp (n = 6).

(D) Levels of 5-HT released from RIN14B cells after exposure to metabolites: acetate (1 mM), α-tocopherol (8 μM), arabinose (50 μM), azelate (50 μM), butyrate (100 μM), cholate (75 μM), deoxycholate (25 μM), ferulate (25 μM), GABA (25 μM), glycine (50 μM), N-methyl proline (0.5 μM), oleanolate (50 μM), p-aminobenzoate (1 μM), propionate (100 μM), taurine (50 μM), and tyramine (100 μM). Data are normalized to 5-HT levels in vehicle-treated controls (gray line at 1) (n = 5–19).

(E) Expression of *TPH1* relative to *GAPDH* in RIN14B cells after metabolite exposure. Data are normalized to expression in vehicle-treated controls (gray line at 1) (n = 3–4).

(F) Levels of 5-HT in colons (left) and serum (center) of GF mice at 30 min after intrarectal injection of deoxycholate (125 mg/kg) or vehicle. Expression of *TPH1* relative to *GAPDH* (right) at 1 hr post injection (n = 3–8).

(legend continued on next page)

reported to modulate the microbiota (Islam et al., 2011) and the severity of *Clostridium difficile* and *Camphylobacter jejuni* infections (Buffie et al., 2014; Malik-Kale et al., 2008). Detrimental effects are also observed; deoxycholate exhibits carcinogenic properties and is linked to various cancers (Bernstein et al., 2011; Yoshimoto et al., 2013). Notably, deoxycholate is reported to promote GI motility by activating TGR5 G protein-coupled receptors on ECs (Alemi et al., 2013), which is consistent with our finding that Sp-induced metabolites raise 5-HT levels in ECs and that Sp colonization improves GI motility. Particular *Clostridium* species are known to possess high 7 $\alpha$ -dehydroxylation activity required for the production of deoxycholate from cholate (Kitahara et al., 2001; Narushima et al., 2006), which is in line with our finding that Sp microbes, comprised largely of Clostridia, increase deoxycholate levels. Deoxycholate concentrations are substantially higher in the colon versus small intestine (Sayin et al., 2013), which, coupled to the finding that bacterial load and diversity is greater in the colon versus small intestine (Sekirov et al., 2010), could contribute to the regional specificity of microbiota-mediated increases in 5-HT synthesis to colonic ECs. Phylogenetic analysis of 16S rDNA sequences reveals that a subset of microbes recovered from Sp-colonized mice cluster taxonomically with known 7 $\alpha$ -dehydroxylating Clostridia (Figures 6G and S7). Notably, there are striking phylogenetic commonalities between taxa identified in Sp- and hSp-colonized mice (Figure S7), consistent with their very similar luminal metabolomic profiles (Figure 6C) and ability to promote 5-HT synthesis from colonic ECs (Figure S3).

We also reveal that the metabolites  $\alpha$ -tocopherol, tyramine, and PABA are elevated in feces by Sp. hSp or SPF colonization co-vary with fecal 5-HT levels and sufficiently induce 5-HT in vitro and in vivo (Figures 6 and S6; Table S1).  $\alpha$ -tocopherol is a naturally abundant form of vitamin E, with reported therapeutic effects for several diseases (Brigelius-Flohe and Traber, 1999). Interestingly, patients with depression exhibit decreased plasma  $\alpha$ -tocopherol (Maes et al., 2000; Owen et al., 2005), and treatment with  $\alpha$ -tocopherol reduces depressive-like behavior in pre-clinical models (Lobato et al., 2010), suggesting a link between  $\alpha$ -tocopherol and 5-HT-related disease. Tyramine is a trace amine that acts as a neurotransmitter and catecholamine-releasing agent. Particular bacteria can produce tyramine by decarboxylation of tyrosine in the gut, where tyramine is reported to stimulate fast ileal contractions and neuropeptide Y release (Marcobal et al., 2012). PABA is an intermediate of folic acid synthesis and essential nutrient for some bacteria. Particular species can generate PABA from chorismate (de Crécy-Lagard et al., 2007), but physiological roles for PABA in the GI tract are unclear. Subsets of microbes from Sp- and hSp-colonized mice relate phylogenetically to Clostridia with putative genes for  $\alpha$ -tocopherol and tyrosine metabolism (Figures 6G and S7). Screening Sp microbes for target metabolic functions could

serve as a tractable approach for further parsing the Sp consortium into the minimal species required for increasing 5-HT biosynthesis by ECs.

While there is increasing evidence for a bi-directional relationship between the gut microbiota and gut sensorimotor function, the particular microbes and mechanisms involved are unclear. The microbiota is required for normal IPAN excitability (McVey Neufeld et al., 2013), and recent studies reveal that changes in the microbiota can alter levels of neuroactive molecules, such as nitric oxide, substance P and endocannabinoids, which have the potential to influence gut motor activity (Quigley, 2011). Mucosal immune responses (Collins, 1996), including key interactions between macrophages and enteric neurons (Muller et al., 2014), also modulate GI motility via the gut microbiota. It will be interesting to determine whether 5-HT-mediated effects on immunity (Baganz and Blakely, 2013) contribute to its effects on GI motility. Notably, deconjugated bile salts are reported to alter gut sensorimotor activity (Appleby and Walters, 2014), which supports our hypothesis that Sp-induced increases in deoxycholate, among other metabolites, contribute to its ability to elevate colonic 5-HT and decrease intestinal transit time.

While we demonstrate that Sp-mediated induction of colonic and blood 5-HT regulates GI motility and platelet function in mice, further research is needed to explore additional implications of microbially induced 5-HT on host health and disease (O'Mahony et al., 2015). Peripheral 5-HT modulates several cellular processes, including osteoblast differentiation, erythropoiesis and immunity. Moreover, gross abnormalities in brain structure are observed in Tph1<sup>+/-</sup> embryos from Tph1<sup>-/-</sup> mothers (Côté et al., 2007), indicating that maternal peripheral 5-HT is important for offspring neurodevelopment. Placentally-derived 5-HT also influences neurodevelopment, influencing thalamocortical axon guidance (Bonnin et al., 2011). Interestingly, the indigenous microbiota also modulates hippocampal levels of 5-HT (Clarke et al., 2013), revealing a role for the microbiota in regulating the brain serotonergic system. Overall, our findings provide a mechanism by which select microbes and their metabolic products can be used to promote endogenous, localized 5-HT biosynthesis and further alter host physiology.

## EXPERIMENTAL PROCEDURES

See Supplemental Information for additional details and references.

### PCPA Treatment

At 2 weeks post-bacterial treatment, mice were anesthetized with isoflurane, and PCPA (90 mg/kg) (Liu et al., 2008) was administered intrarectally every 12 hr for 3 days using a sterile 3.5 Fr silicone catheter inserted 4 cm into the rectum. Mice were suspended by tail for 30 s before return to the home

(G) Phylogenetic tree displaying key Sp. (M) and hSp. (H) operational taxonomic units (OTUs) relative to reference *Clostridium* species with reported 7 $\alpha$ -dehydroxylation activity (red circles). Relative abundance is indicated in parentheses (n = 3).

Data are presented as mean  $\pm$  SEM. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001. n.s., not statistically significant; SPF, specific pathogen-free (conventionally-colonized); GF, germ-free; Sp, spore-forming bacteria; iono, 15 uM ionomycin; ASF, Altered Schaedler Flora; hSp, human-derived spore-forming bacteria.

See also Figures S6 and S7.

cage. For mock treatment, mice were anesthetized and intrarectally injected with sterile water as vehicle.

### Serotonin Measurements

Serotonin levels were detected in sera and supernatant of tissue homogenates by ELISA according to the manufacturer's instructions (Eagle Biosciences). Readings from tissue samples were normalized to total protein content as detected by BCA assay (Thermo Pierce). Data compiled across multiple experiments are expressed as 5-HT concentrations normalized to SPF controls within each experiment.

### RIN14B In Vitro Culture Experiments

RIN14B cells (ATCC) were seeded at  $10^5$  cells/cm<sup>2</sup> and cultured according to methods described in Nozawa et al. (2009). Total colonic luminal contents were collected from adult SPF, GF, and GF mice colonized with spore-forming bacteria, suspended at 120  $\mu$ l/mg in HBSS supplemented with 0.1% BSA and 2  $\mu$ M fluoxetine, and centrifuged at  $12,000 \times g$  for 10 min. Supernatants were passed through 0.2  $\mu$ m pore syringe filters. Cultured RIN14B cells were incubated with colonic luminal filtrate for 1 hr at 37°C.

### GI Transit Assay

Mice were orally gavaged with 200  $\mu$ l sterile solution of 6% carmine red (Sigma Aldrich) and 0.5% methylcellulose (Sigma Aldrich) in water and placed in a new cage with no bedding (Li et al., 2011). Starting at 120 min post-gavage, mice were monitored every 10 min for production of a red fecal pellet. GI transit time was recorded as the total number of minutes elapsed (rounded to the nearest 10 min) before production of a red fecal pellet. For mice treated intrarectally with PCPA or metabolites, GI transit assay was conducted 1 hr after the third injection.

### Platelet Activation and Aggregation Assays

Blood samples were collected by cardiac puncture, diluted with a 2 $\times$  volume of HEPES medium and centrifuged through PST lithium heparin vacutainers (Becton Dickinson). Expression of platelet activation markers was measured by flow cytometry (Nieswandt et al., 2004; Ziu et al., 2012). Platelet aggregation assays were conducted according to methods described in (De Cuyper et al., 2013). Remaining unstained PRP was used to generate PRP smears. Slides were stained with Wright Stain (Camco) according to standard procedures.

### 16S rRNA Gene Sequencing and Analysis

Fecal samples were collected at 2 weeks after orally gavaging GF mice with Sp or hSp. Bacterial genomic DNA was extracted from mouse fecal pellets using the QIAamp DNA Stool Mini Kit (QIAGEN). The library was generated according to methods from (Caporaso et al., 2011). The V4 regions of the 16S rRNA gene were PCR amplified, purified and then sequenced using the Illumina MiSeq platform. Operational taxonomic units (OTUs) were chosen de novo with UPARSE pipeline (Edgar, 2013). Taxonomy assignment and rarefaction were performed using QIIME1.8.0 (Caporaso et al., 2010). Phylogenetic trees were built using PhyML (Guindon et al., 2010) and visualized using iTOL (Letunic and Bork, 2007).

See also [Extended Experimental Procedures](#).

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Discussion, Extended Experimental Procedures, seven figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.047>.

### AUTHORS CONTRIBUTIONS

J.M.Y., K.Y., G.P.D., G.G.S., P.A., L.M., and E.Y.H. performed the experiments and analyzed the data. J.M.Y. and E.Y.H. designed the study. C.R.N., R.F.I., and S.K.M. provided novel reagents. R.F.I. and S.K.M. provided valuable support and contributed equally. J.M.Y. and E.Y.H. wrote the manuscript. All authors discussed the results and commented on the manuscript.

### ACKNOWLEDGMENTS

The authors acknowledge the assistance of Andrew Stefka and Taylor Feehley (University of Chicago) for contributing pilot serum and fecal samples, Taren Thron, Sara McBride, and Alyssa Maskell for caring for the animals, Drs. Nathan Dalleska and Jesse Allen (Caltech) for conducting pilot LC/MS experiments, Said Bogatyrev (Caltech) for helpful advice, Natasha Shelby (Caltech) for editing the manuscript, and the late Dr. Paul H. Patterson for his valuable support. This work was supported by the NIH Director's Early Independence Award (5DP5OD017924 to E.Y.H.), Caltech Center for Environmental Microbial Interactions Award (to E.Y.H.), National Science Foundation (NSF) Emerging Frontiers in Research and Innovation Award (EFRI-1137089 to R.F.I. and S.K.M.), National Human Genome Research Institute (NHGRI) grant (R01HG005826 to R.F.I.), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) grant (DK078938 to S.K.M.) and National Institute of Mental Health (NIMH) grant (MH100556 to S.K.M.), National Institute of Allergy and Infectious Diseases (NIAID) grant (AI106302 to C.R.N.), and Food Allergy Research and Education (FARE) and University of Chicago Digestive Diseases Center Core Grant (P30DK42086 to C.R.N.).

Received: September 25, 2014

Revised: December 16, 2014

Accepted: February 18, 2015

Published: April 9, 2015

### REFERENCES

- Abdala-Valencia, H., Berdnikovs, S., McCary, C.A., Urick, D., Mahadevia, R., Marchese, M.E., Swartz, K., Wright, L., Mutlu, G.M., and Cook-Mills, J.M. (2012). Inhibition of allergic inflammation by supplementation with 5-hydroxytryptophan. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 303, L642–L660.
- Alemi, F., Poole, D.P., Chiu, J., Schoonjans, K., Cattaruzza, F., Grider, J.R., Bunnett, N.W., and Corvera, C.U. (2013). The receptor TGR5 mediates the prokinetic actions of intestinal bile acids and is required for normal defecation in mice. *Gastroenterology* 144, 145–154.
- Amireault, P., Sibon, D., and Côté, F. (2013). Life without peripheral serotonin: insights from tryptophan hydroxylase 1 knockout mice reveal the existence of paracrine/autocrine serotonergic networks. *ACS Chem. Neurosci.* 4, 64–71.
- Appleby, R.N., and Walters, J.R. (2014). The role of bile acids in functional GI disorders. *J. Neurogastroenterol. Motil.* 26, 1057–1069.
- Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H., Fukuda, S., Saito, T., Narushima, S., Hase, K., et al. (2013). Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* 500, 232–236.
- Baganz, N.L., and Blakely, R.D. (2013). A dialogue between the immune system and brain, spoken in the language of serotonin. *ACS Chem. Neurosci.* 4, 48–63.
- Bernstein, C., Holubec, H., Bhattacharyya, A.K., Nguyen, H., Payne, C.M., Zaitlin, B., and Bernstein, H. (2011). Carcinogenicity of deoxycholate, a secondary bile acid. *Arch. Toxicol.* 85, 863–871.
- Bonnin, A., Goeden, N., Chen, K., Wilson, M.L., King, J., Shih, J.C., Blakely, R.D., Deneris, E.S., and Levitt, P. (2011). A transient placental source of serotonin for the fetal forebrain. *Nature* 472, 347–350.
- Brigelius-Flohe, R., and Traber, M.G. (1999). Vitamin E: function and metabolism. *FASEB J.* 13, 1145–1155.
- Buffie, C.G., Bucci, V., Stein, R.R., McKenney, P.T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., et al. (2014). Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* 517, 205–208.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336.

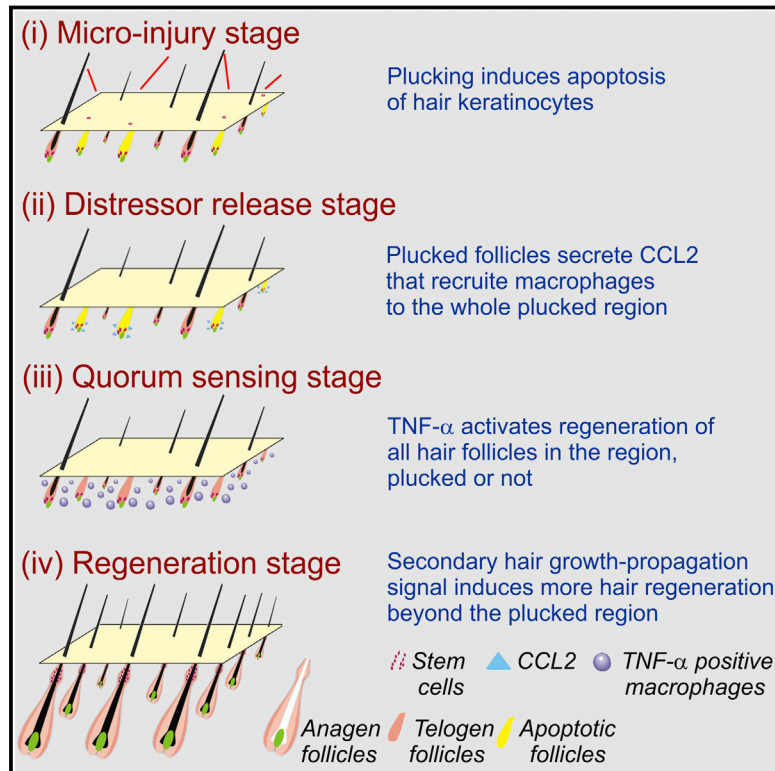
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **108** (Suppl 1), 4516–4522.
- Chabbi-Achengli, Y., Coudert, A.E., Callebort, J., Geoffroy, V., Côté, F., Collet, C., and de Vernejoul, M.C. (2012). Decreased osteoclastogenesis in serotonin-deficient mice. *Proc. Natl. Acad. Sci. USA* **109**, 2567–2572.
- Clarke, G., Grenham, S., Scully, P., Fitzgerald, P., Moloney, R.D., Shanahan, F., Dinan, T.G., and Cryan, J.F. (2013). The microbiome-gut-brain axis during early life regulates the hippocampal serotonergic system in a sex-dependent manner. *Mol. Psychiatry* **18**, 666–673.
- Collins, S.M. (1996). The immunomodulation of enteric neuromuscular function: implications for motility and inflammatory disorders. *Gastroenterology* **111**, 1683–1699.
- Côté, F., Thévenot, E., Fligny, C., Fromes, Y., Darmon, M., Ripoché, M.A., Bayard, E., Hanoun, N., Saurini, F., Lechat, P., et al. (2003). Disruption of the non-neuronal tph1 gene demonstrates the importance of peripheral serotonin in cardiac function. *Proc. Natl. Acad. Sci. USA* **100**, 13525–13530.
- Côté, F., Fligny, C., Bayard, E., Launay, J.M., Gershon, M.D., Mallet, J., and Vojdani, G. (2007). Maternal serotonin is crucial for murine embryonic development. *Proc. Natl. Acad. Sci. USA* **104**, 329–334.
- de Crécy-Lagard, V., El Yacoubi, B., de la Garza, R.D., Noiri, A., and Hanson, A.D. (2007). Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. *BMC Genomics* **8**, 245.
- De Cuyper, I.M., Meinders, M., van de Vijver, E., de Korte, D., Porcelijn, L., de Haas, M., Eble, J.A., Seeger, K., Rutella, S., Pagliara, D., et al. (2013). A novel flow cytometry-based platelet aggregation assay. *Blood* **121**, e70–e80.
- Dewhirst, F.E., Chien, C.C., Paster, B.J., Ericson, R.L., Orcutt, R.P., Schauer, D.B., and Fox, J.G. (1999). Phylogeny of the defined murine microbiota: altered Schaedler flora. *Appl. Environ. Microbiol.* **65**, 3287–3292.
- Ducy, P., and Karsenty, G. (2010). The two faces of serotonin in bone biology. *J. Cell Biol.* **191**, 7–13.
- Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998.
- Fukumoto, S., Tatewaki, M., Yamada, T., Fujimiyama, M., Mantyh, C., Voss, M., Eubanks, S., Harris, M., Pappas, T.N., and Takahashi, T. (2003). Short-chain fatty acids stimulate colonic transit via intraluminal 5-HT release in rats. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **284**, R1269–R1276.
- Gershon, M.D., and Tack, J. (2007). The serotonin signaling system: from basic understanding to drug development for functional GI disorders. *Gastroenterology* **132**, 397–414.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321.
- Hirano, S., Nakama, R., Tamaki, M., Masuda, N., and Oda, H. (1981). Isolation and characterization of thirteen intestinal microorganisms capable of 7  $\alpha$ -dehydroxylating bile acids. *Appl. Environ. Microbiol.* **41**, 737–745.
- Hoffman, J.M., Tyler, K., MacEachern, S.J., Balemba, O.B., Johnson, A.C., Brooks, E.M., Zhao, H., Swain, G.M., Moses, P.L., Galligan, J.J., et al. (2012). Activation of colonic mucosal 5-HT(4) receptors accelerates propulsive motility and inhibits visceral hypersensitivity. *Gastroenterology* **142**, 844–854.
- Islam, K.B., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., Ogura, Y., Hayashi, T., and Yokota, A. (2011). Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology* **141**, 1773–1781.
- Kitahara, M., Takamine, F., Imamura, T., and Benno, Y. (2000). Assignment of *Eubacterium* sp. VPI 12708 and related strains with high bile acid 7 $\alpha$ -dehydroxylating activity to *Clostridium* *scindens* and proposal of *Clostridium* *hyemonae* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **50**, 971–978.
- Kitahara, M., Takamine, F., Imamura, T., and Benno, Y. (2001). *Clostridium* *hiranonis* sp. nov., a human intestinal bacterium with bile acid 7 $\alpha$ -dehydroxylating activity. *Int. J. Syst. Evol. Microbiol.* **51**, 39–44.
- Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128.
- Li, Z., Chalazonitis, A., Huang, Y.Y., Mann, J.J., Margolis, K.G., Yang, Q.M., Kim, D.O., Cote, F., Mallet, J., and Gershon, M.D. (2011). Essential roles of enteric neuronal serotonin in gastrointestinal motility and the development/survival of enteric dopaminergic neurons. *J. Neurosci.* **31**, 8998–9009.
- Liu, Q., Yang, Q., Sun, W., Vogel, P., Heydorn, W., Yu, X.Q., Hu, Z., Yu, W., Jonas, B., Pineda, R., et al. (2008). Discovery and characterization of novel tryptophan hydroxylase inhibitors that selectively inhibit serotonin synthesis in the gastrointestinal tract. *J. Pharmacol. Exp. Ther.* **325**, 47–55.
- Liu, Y., Jennings, N.L., Dart, A.M., and Du, X.J. (2012). Standardizing a simpler, more sensitive and accurate tail bleeding assay in mice. *World J. Exp. Med.* **2**, 30–36.
- Lobato, K.R., Cardoso, C.C., Binfaré, R.W., Budni, J., Wagner, C.L., Brocardo, P.S., de Souza, L.F., Brocardo, C., Flesch, S., Freitas, A.E., et al. (2010).  $\alpha$ -Tocopherol administration produces an antidepressant-like effect in predictive animal models of depression. *Behav. Brain Res.* **209**, 249–259.
- Ma, L., Kim, J., Hatzenpichler, R., Karymov, M.A., Hubert, N., Hanan, I.M., Chang, E.B., and Ismagilov, R.F. (2014). Gene-targeted microfluidic cultivation validated by isolation of a gut bacterium listed in Human Microbiome Project's Most Wanted taxa. *Proc. Natl. Acad. Sci. USA* **111**, 9768–9773.
- Maes, M., De Vos, N., Pioli, R., Demedts, P., Wauters, A., Neels, H., and Christophe, A. (2000). Lower serum vitamin E concentrations in major depression. Another marker of lowered antioxidant defenses in that illness. *J. Affect. Disord.* **58**, 241–246.
- Malik-Kale, P., Parker, C.T., and Konkel, M.E. (2008). Culture of *Campylobacter jejuni* with sodium deoxycholate induces virulence gene expression. *J. Bacteriol.* **190**, 2286–2297.
- Marcobal, A., De las Rivas, B., Landete, J.M., Tabera, L., and Muñoz, R. (2012). Tyramine and phenylethylamine biosynthesis by food bacteria. *Crit. Rev. Food Sci. Nutr.* **52**, 448–467.
- Markowitz, V.M., Chen, I.M., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., et al. (2012). IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122.
- Mawe, G.M., and Hoffman, J.M. (2013). Serotonin signalling in the gut—functions, dysfunctions and therapeutic targets. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 473–486.
- McVey Neufeld, K.A., Mao, Y.K., Bienenstock, J., Foster, J.A., and Kunze, W.A. (2013). The microbiome is essential for normal gut intrinsic primary afferent neuron excitability in the mouse. *Neurogastroenterol. Motil.* **25**, e183–e188.
- Mercado, C.P., Quintero, M.V., Li, Y., Singh, P., Byrd, A.K., Talabnin, K., Ishihara, M., Azadi, P., Rusch, N.J., Kuberan, B., et al. (2013). A serotonin-induced N-glycan switch regulates platelet aggregation. *Sci. Rep.* **3**, 2795.
- Muller, P.A., Koscsó, B., Rajani, G.M., Stevanovic, K., Berres, M.L., Hashimoto, D., Mortha, A., Leboeuf, M., Li, X.M., Mucida, D., et al. (2014). Crosstalk between muscularis macrophages and enteric neurons regulates gastrointestinal motility. *Cell* **158**, 300–313.
- Narushima, S., Itoha, K., Miyamoto, Y., Park, S.H., Nagata, K., Kuruma, K., and Uchida, K. (2006). Deoxycholic acid formation in gnotobiotic mice associated with human intestinal bacteria. *Lipids* **41**, 835–843.
- Nieswandt, B., Schulte, V., and Bergmeier, W. (2004). Flow-cytometric analysis of mouse platelet function. *Methods Mol. Biol.* **272**, 255–268.
- Nozawa, K., Kawabata-Shoda, E., Doihara, H., Kojima, R., Okada, H., Mochizuki, S., Sano, Y., Inamura, K., Matsushime, H., Koizumi, T., et al. (2009). TRPA1 regulates gastrointestinal motility through serotonin release from enterochromaffin cells. *Proc. Natl. Acad. Sci. USA* **106**, 3408–3413.



- O'Mahony, S.M., Clarke, G., Borre, Y.E., Dinan, T.G., and Cryan, J.F. (2015). Serotonin, tryptophan metabolism and the brain-gut-microbiome axis. *Behav. Brain Res.* 277, 32–48.
- Odell, T.T., Jr., and McDonald, T.P. (1961). Life span of mouse blood platelets. *Proc. Soc. Exp. Biol. Med.* 106, 107–108.
- Oleskin, A.V., Kirovskaja, T.A., Botvinko, I.V., and Lysak, L.V. (1998). [Effect of serotonin (5-hydroxytryptamine) on the growth and differentiation of microorganisms]. *Mikrobiologiya* 67, 305–312.
- Owen, A.J., Batterham, M.J., Probst, Y.C., Grenyer, B.F., and Tapsell, L.C. (2005). Low plasma vitamin E levels in major depression: diet or disease? *Eur. J. Clin. Nutr.* 59, 304–306.
- Quigley, E.M. (2011). Microflora modulation of motility. *J. Neurogastroenterol. Motil.* 17, 140–147.
- Ramage, A.G., and Villalón, C.M. (2008). 5-hydroxytryptamine and cardiovascular regulation. *Trends Pharmacol. Sci.* 29, 472–481.
- Reikvam, D.H., Erofeev, A., Sandvik, A., Grcic, V., Jahnsen, F.L., Gaustad, P., McCoy, K.D., Macpherson, A.J., Meza-Zepeda, L.A., and Johansen, F.E. (2011). Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PLoS ONE* 6, e17996.
- Roshchina, V.V. (2010). Evolutionary considerations of neurotransmitters in microbial, plant, and animal cells. *Microbial Endocrinol.*, 17–52.
- Savelieva, K.V., Zhao, S., Pogorelov, V.M., Rajan, I., Yang, Q., Cullinan, E., and Lanthorn, T.H. (2008). Genetic disruption of both tryptophan hydroxylase genes dramatically reduces serotonin and affects behavior in models sensitive to antidepressants. *PLoS ONE* 3, e3301.
- Sayin, S.I., Wahlström, A., Felin, J., Jäntti, S., Marschall, H.U., Bamberg, K., Angelin, B., Hyötyläinen, T., Orešić, M., and Bäckhed, F. (2013). Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell Metab.* 17, 225–235.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.
- Sekirov, I., Russell, S.I., Antunes, L.C., and Finlay, B.B. (2010). Gut microbiota in health and disease. *Physiol. Rev.* 90, 859–904.
- Shimizu, T., Shimizu, S., Hattori, R., Gabazza, E.C., and Majima, Y. (2003). In vivo and in vitro effects of macrolide antibiotics on mucus secretion in airway epithelial cells. *Am. J. Respir. Crit. Care Med.* 168, 581–587.
- Sjögren, K., Engdahl, C., Henning, P., Lerner, U.H., Tremaroli, V., Lagerquist, M.K., Bäckhed, F., and Ohlsson, C. (2012). The gut microbiota regulates bone mass in mice. *J. Bone Miner. Res.* 27, 1357–1367.
- Stasi, C., Bellini, M., Bassotti, G., Blandizzi, C., and Milani, S. (2014). Serotonin receptors and their role in the pathophysiology and therapy of irritable bowel syndrome. *Tech. Coloproctol.* 18, 613–621.
- Stefka, A.T., Feehley, T., Tripathi, P., Qiu, J., McCoy, K., Mazmanian, S.K., Tjota, M.Y., Seo, G.Y., Cao, S., Theriault, B.R., et al. (2014). Commensal bacteria protect against food allergen sensitization. *Proc. Natl. Acad. Sci. USA* 111, 13145–13150.
- Tsavkelova, E.A., Klimova, S.I., Cherdyntseva, T.A., and Netrusov, A.I. (2006). [Hormones and hormone-like substances of microorganisms: a review]. *Prikl. Biokhim. Mikrobiol.* 42, 261–268.
- Uribe, A., Alam, M., Johansson, O., Midtvedt, T., and Theodorsson, E. (1994). Microflora modulates endocrine cells in the gastrointestinal mucosa of the rat. *Gastroenterology* 107, 1259–1269.
- Wade, P.R., Chen, J., Jaffe, B., Kassem, I.S., Blakely, R.D., and Gershon, M.D. (1996). Localization and function of a 5-HT transporter in crypt epithelia of the gastrointestinal tract. *J. Neurosci.* 16, 2352–2364.
- Walther, D.J., Peter, J.U., Bashammakh, S., Hörtnagl, H., Voits, M., Fink, H., and Bader, M. (2003). Synthesis of serotonin by a second tryptophan hydroxylase isoform. *Science* 299, 76.
- Westphal, J.F., Vetter, D., and Brogard, J.M. (1994). Hepatic side-effects of antibiotics. *J. Antimicrob. Chemother.* 33, 387–401.
- Wikoff, W.R., Anfora, A.T., Liu, J., Schultz, P.G., Lesley, S.A., Peters, E.C., and Siuzdak, G. (2009). Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. USA* 106, 3698–3703.
- Williams, B.B., Van Benschoten, A.H., Cimermanic, P., Donia, M.S., Zimmermann, M., Taketani, M., Ishihara, A., Kashyap, P.C., Fraser, J.S., and Fischbach, M.A. (2014). Discovery and characterization of gut microbiota decarboxylases that can produce the neurotransmitter tryptamine. *Cell Host Microbe* 16, 495–503.
- Yadav, V.K., Ryu, J.H., Suda, N., Tanaka, K.F., Gingrich, J.A., Schütz, G., Gorieux, F.H., Chiang, C.Y., Zajac, J.D., Insogna, K.L., et al. (2008). Lrp5 controls bone formation by inhibiting serotonin synthesis in the duodenum. *Cell* 135, 825–837.
- Yoshimoto, S., Loo, T.M., Atarashi, K., Kanda, H., Sato, S., Oyadomari, S., Iwakura, Y., Oshima, K., Morita, H., Hattori, M., et al. (2013). Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* 499, 97–101.
- Zhao, Y., Lee, M.J., Cheung, C., Ju, J.H., Chen, Y.K., Liu, B., Hu, L.Q., and Yang, C.S. (2010). Analysis of multiple metabolites of tocopherols and tocotrienols in mice and humans. *J. Agric. Food Chem.* 58, 4844–4852.
- Ziu, E., Mercado, C.P., Li, Y., Singh, P., Ahmed, B.A., Freyaldenhoven, S., Lensing, S., Ware, J., and Kilic, F. (2012). Down-regulation of the serotonin transporter in hyperreactive platelets counteracts the pro-thrombotic effect of serotonin. *J. Mol. Cell. Cardiol.* 52, 1112–1121.

# Organ-Level Quorum Sensing Directs Regeneration in Hair Stem Cell Populations

## Graphical Abstract



## Authors

Chih-Chiang Chen, Lei Wang, ..., Arthur D. Lander, Cheng Ming Chuong

## Correspondence

cmchuong@usc.edu

## In Brief

Organ-level quorum sensing through a multi-step immune cascade allows collective regeneration of hair follicle populations. This mechanism provides a means to assess the magnitude and extent of injury that the skin has sustained and make an all-or-none decision whether to regenerate.

## Highlights

- Quorum sensing underlies collective regenerative behavior in a hair follicle population
- Sensing occurs via injury  $\rightarrow$  CCL2  $\rightarrow$  macrophage  $\rightarrow$  TNF- $\alpha$   $\rightarrow$  hair regeneration pathway
- Coupling molecular diffusion and cell mobility achieves a long signaling length scale
- Stem cell social behavior can be exploited to enhance the reliability of regeneration

## Accession Numbers

GSE46181



# Organ-Level Quorum Sensing Directs Regeneration in Hair Stem Cell Populations

Chih-Chiang Chen,<sup>1,2,3</sup> Lei Wang,<sup>4</sup> Maksim V. Plikus,<sup>5,6,7</sup> Ting Xin Jiang,<sup>1</sup> Philip J. Murray,<sup>8</sup> Raul Ramos,<sup>5,6,7</sup> Christian F. Guerrero-Juarez,<sup>5,6,7</sup> Michael W. Hughes,<sup>9</sup> Oscar K. Lee,<sup>10</sup> Songtao Shi,<sup>11</sup> Randall B. Widelitz,<sup>1</sup> Arthur D. Lander,<sup>5,7,12</sup> and Cheng Ming Chuong<sup>1,3,13,\*</sup>

<sup>1</sup>Department of Pathology, University of Southern California, Los Angeles, CA 90033, USA

<sup>2</sup>Department of Dermatology, Taipei Veterans General Hospital, Taipei 112, Taiwan, ROC

<sup>3</sup>Institute of Clinical Medicine and Department of Dermatology, National Yang-Ming University, Taipei, Taiwan 112, ROC

<sup>4</sup>State Key Laboratory of Military Stomatology, Department of Oral and Maxillofacial Surgery, School of Stomatology, The Fourth Military Medical University, Xi'an, Shaanxi 710032, China

<sup>5</sup>Department of Developmental and Cell Biology

<sup>6</sup>Sue and Bill Gross Stem Cell Research Center

<sup>7</sup>Center for Complex Biological Systems

University of California Irvine, Irvine, CA 92697, USA

<sup>8</sup>Division of Mathematics, University of Dundee, Dundee DD1 4HN, UK

<sup>9</sup>International Laboratory of Wound Repair and Regeneration and Institute of Clinical Medicine, National Cheng Kung University, Tainan 701, Taiwan, ROC

<sup>10</sup>Department of Orthopaedics and Traumatology, Taipei Veterans General Hospital, Taipei and Center for Stem Cell Research, National Yang-Ming University and Veterans General Hospital, Taipei 112, Taiwan, ROC

<sup>11</sup>Department of Anatomy and Cell Biology, University of Pennsylvania, School of Dental Medicine, Philadelphia, PA 19104, USA

<sup>12</sup>Department of Biomedical Engineering, University of California Irvine, Irvine, CA 92697, USA

<sup>13</sup>Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei 10617, Taiwan

\*Correspondence: [cmchuong@usc.edu](mailto:cmchuong@usc.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.016>

## SUMMARY

Coordinated organ behavior is crucial for an effective response to environmental stimuli. By studying regeneration of hair follicles in response to patterned hair plucking, we demonstrate that organ-level quorum sensing allows coordinated responses to skin injury. Plucking hair at different densities leads to a regeneration of up to five times more neighboring, unplucked resting hairs, indicating activation of a collective decision-making process. Through data modeling, the range of the quorum signal was estimated to be on the order of 1 mm, greater than expected for a diffusible molecular cue. Molecular and genetic analysis uncovered a two-step mechanism, where release of CCL2 from injured hairs leads to recruitment of TNF- $\alpha$ -secreting macrophages, which accumulate and signal to both plucked and unplucked follicles. By coupling immune response with regeneration, this mechanism allows skin to respond predictively to distress, disregarding mild injury, while meeting stronger injury with full-scale cooperative activation of stem cells.

## INTRODUCTION

The effective coordination of organ behavior, either under physiological conditions or as a response to injury, is essential for survival. Integration at the level of large-scale organ systems has

been extensively studied, but the role of shorter range, local coordination has not. For example, is the regeneration of repeated tissue units within an organ (e.g., hair follicles [HFs] in skin, villi in intestine) coordinated so as to achieve collective decision-making? If so, what are the mechanisms of communication, and how is information integrated? In particular, if injury or malfunction affects only a subset of tissue units in an organ, how is such a collective decision made whether to mount a response that is local (e.g., local repair) or global (e.g., tissue level regeneration)?

Mammalian skin offers an excellent platform to address such questions, because its numerous HFs behave as discrete, repeating, semi-autonomous tissue units (Jahoda and Christiano, 2011) distributed on a 2D plane. HFs undergo cyclic regeneration (Paus et al., 1998) by regulating both intra- and extra-follicular cues for hair stem cell activation (Stenn and Paus, 2001; Plikus et al., 2008, 2011; Festa et al., 2011; Chen and Chuong, 2012), both during physiological regeneration and in response to injury (Chuong et al., 2012). The experimental accessibility of HFs makes them an ideal model to study collective decision making in an organ population in vivo.

Classical studies show that hair plucking produces a micro-injury that can potentially lead to hair regeneration (Collins, 1918; Silver and Chase, 1970). This process is thought to be mediated by an autonomous mechanism in each follicle, in which early apoptosis in the bulge leads to activation of hair germ progenitors (Ito et al., 2002). Here, we uncover evidence that the decision of hair stem cells to be activated or remain quiescent also depends on information coming from neighboring follicles. This possibility was first suggested by our earlier study in which plucking fewer than 50 refractory telogen hairs did not induce hair regeneration, while plucking more than 200 hairs did (Plikus

et al., 2008). Here, by varying the spacing, arrangement, and shapes of plucked regions, we unexpectedly found that plucking 200 hairs, with a proper topological distribution can cause up to 1,200 hairs to regenerate. These results demonstrate marked non-autonomy in HF regeneration and a distinctly non-linear quantitative relationship between plucking and regeneration.

As discussed below, the collective HF response to injury may be seen as an example of quorum sensing, a form of social behavior in which population decisions depend on the density of signaling individuals within a given spatial territory (Bassler, 2002; Pratt, 2005). In order to gain insights into the possible mechanisms underlying this behavior, we first used mathematical modeling to identify the characteristic spatial range over which quorums are sensed, which led us to suspect that the signaling mechanism consists of more than just a diffusible molecule. The time course of molecular changes after plucking, together with the results of genetic and pharmacological manipulation, implicated a two-stage mechanism, involving the release of diffusible signals that recruit immune cells (M1 macrophages) which then actively spread among follicles, where they locally induce regeneration through the release of substances such as  $Tnf-\alpha$ .

This work identifies a mechanism for quorum sensing that operates on the millimeter scale to coordinate the behaviors of semi-autonomous tissue units within an organ. Such coordination enables the skin to condition its responses to the spatial extent of injuries, launching a full scale regenerative response only when a sufficient threshold is reached.

## RESULTS

### Topology-Dependent Hair Plucking Can Induce the Regeneration of More Hairs Than Were Plucked by Activating Neighboring, Unplucked Follicles

To gain insight into the mechanisms leading to hair renewal following follicle injury, the relationship between hair plucking density and regeneration were examined in the mouse. To standardize the experiments, we synchronized all the dorsal pelage HFs into refractory telogen before plucking (see [Extended Experimental Procedures](#)). Normal hair density in adult C57BL/6 mouse dorsal skin is  $\sim 45\text{--}60$  hairs/ $\text{mm}^2$ , corresponding to a distance between each follicle of  $\sim 0.15$  mm ([Figure S1E](#)). In the first set of experiments, 200 evenly distributed refractory telogen hairs were plucked within a circular skin area (the “injury field”; [Figures 1B](#) and [1C](#), red circle). By plucking a constant number of hairs but altering the size of the injury field ([Figures 1A](#), [1B](#), and [S1](#)), plucking densities from 2–50 hairs/ $\text{mm}^2$  were obtained ([Figures 1](#) and [S1](#); [Extended Experimental Procedures](#)). We then studied the regenerative behavior of the HFs.

We observe three types of responses ([Figure 1F](#)). First, if 200 hairs were plucked in a large area ( $>6$  mm diameter,  $28.3\text{ mm}^2$ , plucking density  $<10$  hairs/ $\text{mm}^2$ ) ([Figures 1B](#) and [S1](#)), no regeneration of plucked or unplucked follicles occurs even after 30 days ([Figures 1A](#), [1B](#), and [S1](#)). This is because the plucking density is too low and does not generate accumulated signals above the threshold level ([Figure 1F](#), zone of very low density plucking, gray area). Second, when 200 hairs are plucked from 3-, 4-, or 5-mm diameter circular areas (plucking density  $>10$  hairs/ $\text{mm}^2$ , the threshold density), we induce a simultaneous regeneration

of the whole region (including the plucked and surrounding unplucked follicles) ([Figures 1C](#) and [1D](#)). Thus, by plucking only 200 hairs, the eventual regeneration of approximately 450, 780, or 1,300 hairs are obtained (with 200 hairs plucked in injury field sizes of 3, 4, and 5 mm in diameter, or 7.1, 12.6, and  $19.6\text{ mm}^2$ , respectively) ([Figures 1C–1F](#), [S1](#), and [S2](#)). As an example, we can induce regeneration of up to 600 unplucked hairs within a 5-mm plucked region (zone of quorum sensing-dependent hair regeneration, orange/light green area in [Figure 1F](#)) and 400 hairs outside of the plucked region, resulting from propagation. Third, when 200 hairs are plucked from a 2.4-mm diameter region (high density, 100% plucking), every follicle in the field is plucked ([Figure 1A](#)). In this case, all follicles re-entered anagen  $\sim 12$  days ( $12.3 \pm 3.37$ ,  $n = 13$ ) after plucking, and the number of regenerating follicles equals the number of plucked follicles (zone of all follicles plucked, dark green). Plucking-dependent regeneration from refractory telogen requires the plucking of at least 50 follicles to reach the basal threshold (Plikus et al., 2008). This zone is equivalent to the frequently used wax stripping procedure (Müller-Röver et al., 2001) in which melted wax was used to strip away all follicles in a large region, usually centimeters in diameter or bigger. This method involves thousands of HFs which will regenerate in synchrony without using quorum sensing.

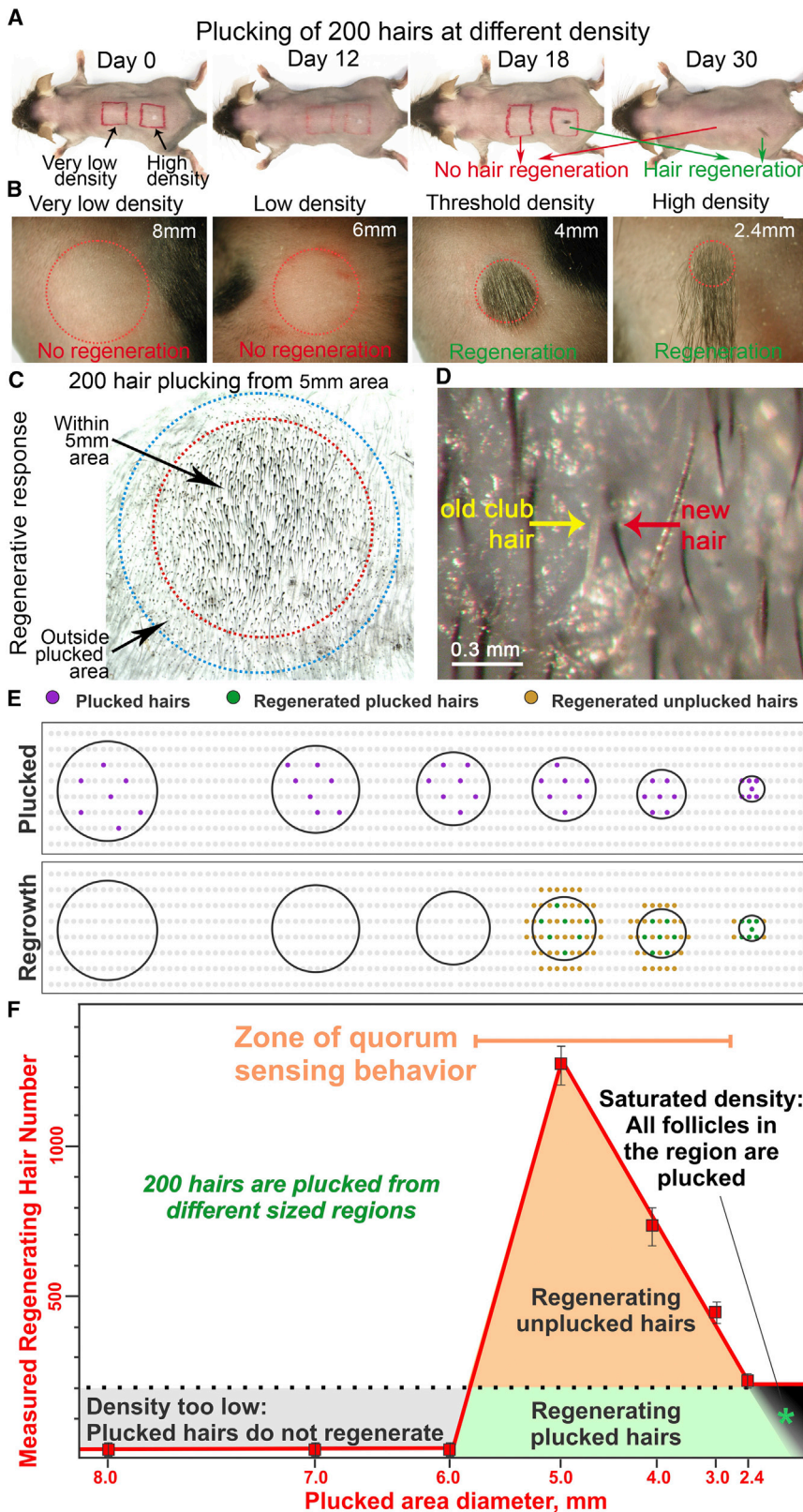
### The Hair Follicle Population as a Quorum-Sensing System

The density-dependence of regeneration, together with the simultaneous regeneration of both plucked and unplucked follicles within the injury field, suggests that plucked follicles produce a signal that (1) spreads to neighboring follicles, (2) accumulates to a level that depends upon the density and position of other plucked follicles, and (3) when present above some threshold level will trigger any follicle—plucked or unplucked—to re-enter anagen.

The idea that HFs produce signals that affect other HFs can be inferred from the coordinated waves of hair cycling that travel across the skin of mice and rabbits (Plikus et al., 2008, 2011). Yet the signals that coordinate such “hair waves” cannot explain the collective regenerative responses seen here, at least not those within the injury field itself. This is because hair waves reflect the ability of follicles in anagen to accelerate the progression of neighboring telogen follicles into anagen, whereas plucking causes injured and uninjured follicles to progress from refractory telogen to regenerate collectively and simultaneously (i.e., regeneration is not driven by neighboring anagen follicles). Just outside the plucked injury fields (e.g., [Figure 1C](#), outside of the red circle), however, the ring of delayed regeneration likely reflects the “hair wave” phenomenon, since follicles in this zone enter anagen only after regeneration in neighboring follicles is well underway. To avoid confusion between initial, collective regeneration and later hair wave spreading, the present study focuses exclusively on early regenerative events.

One way to gain insight into the nature of the quorum signal—that we shall initially call the “distressor”—that coordinates collective regeneration is to characterize its decay length, i.e., the characteristic spatial scale over which the strength of the signal decays. Decay lengths quantify the balance between the rate at which a signal spreads and the rate at which it is destroyed or





**Figure 1. Plucking-Induced Hair Regeneration Is a Population-Based Behavior that Depends on the Density and Distribution of Plucked-Hair Follicles within the Unplucked Follicle Population**

(A and B) Plucking 200 hairs from a circular 2.4 mm in diameter area (100% plucking) leads to hair regeneration 12 days later. Plucking 200 hairs in a 12 mm diameter area (100 mm<sup>2</sup> area; low density plucking) fails to induce follicle regeneration even 30 days later.

(C) Plucking induces regeneration of all follicles (the 200 plucked and 600 unplucked) within the plucked area (red circle, 5 mm in diameter). Unplucked follicles (400 HFs in total) outside the plucked area boundary then regenerate due to hair wave propagation (blue circle).

(D) High power view showing unplucked follicle regeneration: the old gray club hair (yellow) is pushed out by the regenerating black anagen hair (red).

(E) In this schematic drawing, gray dots represent telogen HFs. Black lines encircle exemplary plucked regions. Plucked follicles (purple dots). Regenerating plucked HFs (green dots). Regenerating unplucked HFs (tan dots).

(F) Plot showing the hair regeneration response versus the size of the plucked field. For all different field sizes, 200 hairs are plucked evenly dispersed throughout the field. A regenerative response is observed when 200 hairs are plucked at a density above a threshold (10 hairs/mm<sup>2</sup>), which corresponds to plucking 200 hairs from a 5-mm diameter circular surface area (red line). Three responses represented by different colors (gray, tan, green), are observed (please see text for explanation). The quorum sensing zone is highlighted in orange.

See also [Figures S1](#) and [S2](#).

removed. Diffusible molecules that are captured by high-affinity, cell surface receptors (e.g., morphogens, growth factors, cytokines, and chemokines) tend to have relatively short decay lengths, typically on the order of no more than 100  $\mu\text{m}$  (Teleman and Cohen, 2000; Müller et al., 2012; Sarris et al., 2012; Weber et al., 2013; Shimozone et al., 2013), approximately the same scale as the inter-follicular distance in mouse skin. In contrast, as we show in the next section, the decay length of the putative distressor induced by plucking appears to be substantially larger—on the order of 1 mm, or four to six inter-follicular distances.

### Estimating the Range of Action of the Quorum Signal

Regardless of the physical nature of a signal (i.e., if it spreads from cell to cell via an undirected random walk), its spread can usually be modeled as a diffusion process, and it will display a decay length,  $\lambda$ , equal to the square root of the ratio between its diffusivity (an intrinsic measure of how fast it moves) and the rate constant that characterizes its removal or destruction in the tissue through which it spreads (Lander, 2007) (a physical interpretation of  $\lambda$  is the distance over which the steady-state signal from a point source falls by a factor of  $1-1/e$ , or  $\sim 63\%$ ).

The results of modeling plucked hairs as an array of point sources of a diffusible distressor in a 2D medium (Figure 2A; see also Extended Experimental Procedures) tell us that the expected steady-state distressor concentration should be a function not only of plucking density, but also of injury field size, shape, and  $\lambda$ . For example, with a constant plucking density, concentrations of distressor should rise as a function of field size/ $\lambda$ , leveling off as that ratio gets large (Figure 2A). Assuming that regeneration is triggered when the distressor concentration around an HF exceeds a certain threshold, these results suggest that one could estimate the value of  $\lambda$  from a series of experiments in which plucking density and injury size/shape are both varied.

For example, in Figure 2B, data on whether regeneration in circular injury fields occurred (green dots) or failed (red dots) was tabulated as a function of field radius and plucking density (plotted, in this case, as the inverse of the plucked fraction). Fitting the boundary between positive and negative data to the predictions of the steady-state diffusion model yields estimates of  $\lambda$  between 0.6 and 1.6 mm (Extended Experimental Procedures).

The same model also predicts that, for sufficiently large plucked fields and/or sufficiently high plucking densities, immediate regenerative responses should not be limited to the precise boundaries of the injury field, but should extend a small distance beyond those boundaries (here we refer only to regeneration that occurs at the same time as that within the injury field and not what is triggered significantly later by hair wave propagation). Careful examination of experimental data showed that a small rim of early regeneration indeed occurred just outside of some injury fields. Fitting the sizes of these rims to the model (Figure 2C) yields an independent estimate for  $\lambda = 1$  mm.

Finally, the same diffusion model suggests that  $\lambda$  can also be estimated by holding both plucking density and injury field area constant, but varying the shape of the injury field. To test this prediction, experiments were carried out in which 50 hairs were plucked evenly at a density of every other hair, either in a straight line (Figure 2D), a narrow rectangle (6:1 aspect ratio; Figure 2E) or a square (Figure 2F). Under these distinct topological conditions,

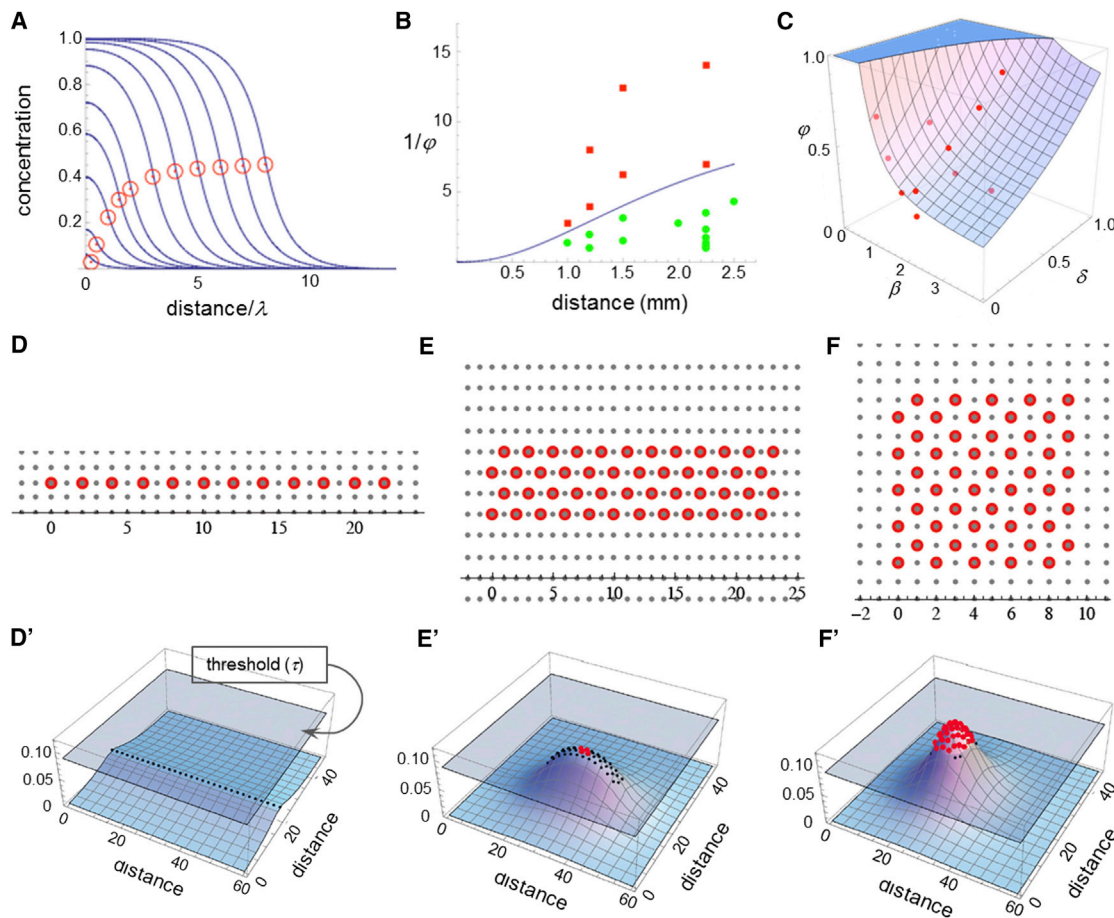
plucked single rows never regenerated, while squares always regenerated robustly. Rectangles occasionally exhibited modest regeneration, suggesting a distressor concentration very close to threshold under these circumstances. Fitting these three behaviors requires a value of  $\lambda$  between 0.7 and 1.2 mm.

The good agreement among these three methods supports the validity of the steady-state diffusion model for describing distressor spreading and places the value of  $\lambda$  at  $\sim 1$  mm. Fitting to the model does not imply, however, that the distressor is a single substance, or even a diffusible molecule, but simply that it spreads according to the same rules. In fact, the observed magnitude of  $\lambda$  suggests that the distressor is not simply a diffusible receptor-binding molecule, since these typically display decay lengths of one tenth this magnitude or less (Teleman and Cohen, 2000; Müller et al., 2012; Sarris et al., 2012; Weber et al., 2013; Shimozone et al., 2013). As described below, further investigation of the molecular nature of the distressor signal supports the idea that it consists of both diffusible molecules and recruited cells that migrate actively between follicles.

### Plucking Induces a Cascade of Inflammatory, Cellular, and Molecular Events

Results from wax-stripping experiments indicate that HF keratinocytes undergo apoptosis  $\sim 4$  hr after injury (Ito et al., 2002; see also Figure 3A). To identify molecules and mechanisms that might be involved in plucking-induced regeneration, we carried out microarray analysis of plucked fields at 12, 24, 48, and 96 hr after injury. Among the notable, time-dependent changes in gene expression, we observed:

- (1) Transient increase in expression of pro-inflammatory cytokines. Immune, inflammatory and wound healing response genes constitute the major portion of early transcriptional activity following plucking. Analyzing the most altered genes by RT-PCR, we found that immune cytokines, chemokine (C-C motif) ligand 2 (CCL2), chemokine (C-X-C motif) ligand 2 (CXCL2), and interleukin 1, beta (IL-1 $\beta$ ) were upregulated soon after plucking (i.e., 12 hr), although expression of these genes peaked at different times (Figure 3B). For example, CCL2 expression peaked around 12 hr (Figure 3B).
- (2) Reduced refractory telogen inhibitor expression. During refractory telogen, the extra-follicular macro-environment expresses high levels of inhibitors, including *Bmp2*, *Dickkopf* (*Dkk1*), and *soluble frizzled related protein* (*Sfrp4*) that block anagen re-entry and hair wave propagation (Plikus et al., 2008, 2011). Expression of *Sfrp4*, a representative gene, decreased markedly at day 1 but rebounded by days 2 and 4 (Figure 3B).
- (3) Increased tumor necrosis factor alpha (*Tnf- $\alpha$* ) expression. *Tnf- $\alpha$*  increases between 1–2 days after plucking and reached a plateau at  $\sim$ day 2 (Figures 3B and 3C). The plateau of *Tnf- $\alpha$*  expression corresponds to a time when activated hairs are in early anagen phase (Figure 3C). We also observe changes of other molecular pathways. For example, *platelet derived growth factor A* (*Pdgf- $\alpha$* ) increased at later stages after plucking (day 4), compatible with published results (Festa et al., 2011) (Figure 3B).



**Figure 2. Mathematical Modeling Identifies the Decay Length of a Putative Quorum Signal**

(A) Calculated steady-state concentrations for a diffusible substance produced within injury fields in proportion to the numbers of plucked HF. Each curve represents a different sized circular injury field, with the red circle placed at the value on the abscissa corresponding to the injury field radius, in units of the diffusing substance decay length. Specifically, the 11 curves represent increasing field sizes of 0.25, 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, and 8 decay lengths. As plucked regions grow larger, the value at the boundary asymptotes to one half the value at the center.  $\lambda$  is a decay length. Please see [Results](#) and [Supplemental Information](#) for more explanation.

(B) Data from a variety of regeneration experiments involving circular wound fields are plotted as a function of the inverse of the plucked fraction ( $\phi$ ) and the radius of the wound field. The curve drawn between the points corresponding to cases of successful (green) and unsuccessful (red) regeneration was obtained from the equations that produced the curves in (A), by fitting two parameters, the decay length and the threshold concentration for regeneration. The range of possible values consistent with the data was manually explored to yield a range of decay length estimates.

(C) The same model was used as in (B), but the data that were fit consisted of the distances,  $\delta$ , just beyond the edges of injury fields at which initial regeneration was seen. ( $\beta$ , radius of the injury field;  $\phi$ , the plucked fraction). The plotted surface represents a least-squares best fit to the data.

(D–F') Effects of injury field shape. Fifty hairs were plucked evenly, at a density of every other hair, either in a straight line (D and D'), a narrow rectangle (6:1 aspect ratio; E and E') or a square (F and F'). In (D')–(F'), a discrete form of the equation used in (A)–(C), in which each HF is modeled as a discrete source, was used to plot the steady-state spatial distributions of a distressor released by plucked follicles (distances are plotted in units of the inter-follicular distance,  $\sim 0.15$  mm). Wherever plotted surfaces extend above a regeneration concentration threshold (gray plane), red dots mark the location of each HF indicating successful regeneration. The requirement that these curves be consistent with the observed regeneration patterns in all three cases was sufficient to provide yet a third estimate of the distressor decay length. See also [Supplemental Information](#) on mathematical model.

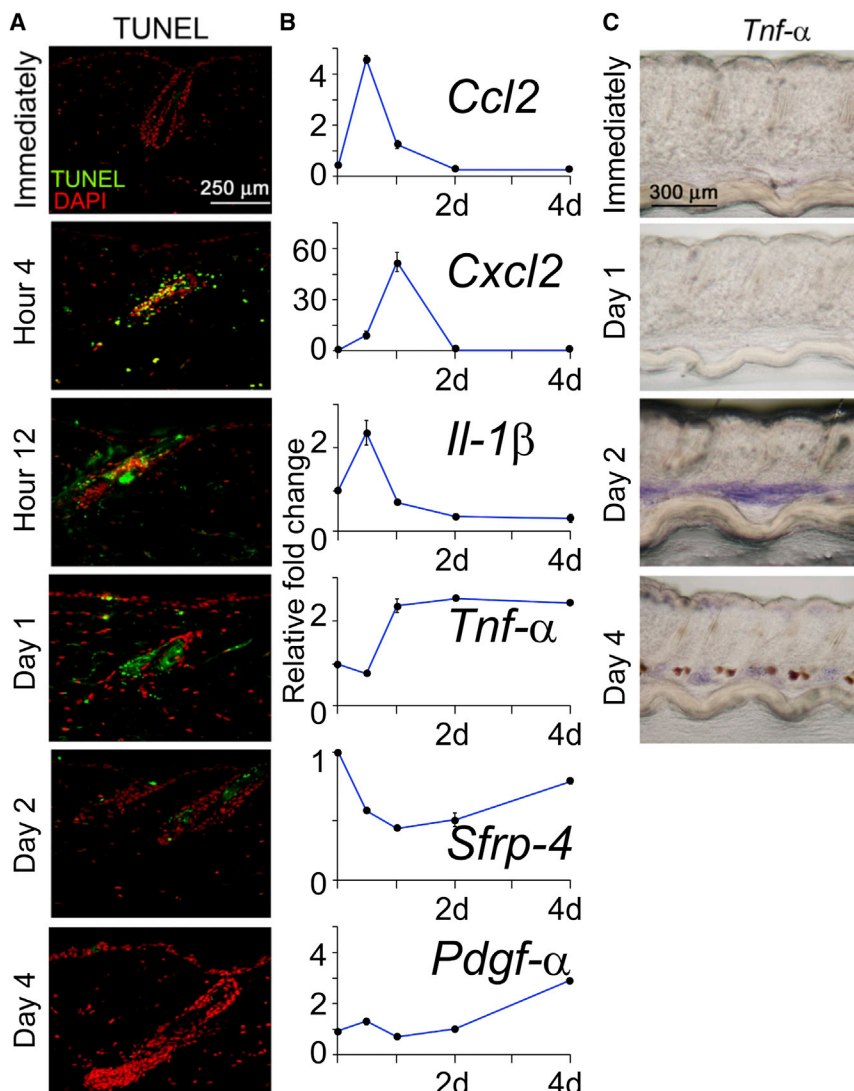
Since the Wnt pathway is critical for hair growth ([Enshell-Seijffers et al., 2010](#); [Lowry et al., 2005](#)), we examined the expression of Wnt pathway members, using whole mount in situ hybridization, and compared their expression patterns, over time, with those of *Tnf- $\alpha$* . *Wnt6*,  $\beta$ -catenin, and *lymphocyte enhancer factor (Lef-1)* were upregulated within new anagen follicles at day 4, but not in the extra-follicular dermal macro-environment ([Figure S3](#)). We also localized *Tnf- $\alpha$*  expression

to the extra-follicular dermal macro-environment ([Figures 3C and S3](#)).

### CCL2 Is a Key Component of the Quorum Signal

The earliest noted signaling molecule expression change that could potentially communicate information from plucked to unplucked follicles was CCL2 ([Figure 3B](#)). Immunohistochemistry showed that CCL2 is primarily produced by HF keratinocytes





**Figure 3. Identification of Macro-Environmental Modulators following Hair Plucking**

(A) TUNEL assay to measure apoptosis. (B) Real-time PCR from extra-follicular macro-environmental tissues revealed the kinetics of gene expression induced by plucking (normalized to GAPDH with 40 cycles, data are represented as mean  $\pm$  SD,  $n = 3$ ). (C) Whole mount in situ hybridization showed that *Tnf-α* is markedly upregulated in the inter-follicular area beginning 2 days after wax stripping. See also Figures S3 and S4.

when the backs of CCL2 null mice were wax-stripped, follicles remained in telogen 3 days after waxing and were still in anagen III to IV at day 6 (Figures 4C and S5A). This delayed hair regrowth in CCL2 null mice following plucking supports a role for CCL2 in plucking induced hair regeneration.

TUNEL staining performed 1 day after plucking revealed that both wild-type and CCL2 null mice showed apoptotic HF cells (Figures 4D and S5B). These results indicate that CCL2 is not required for the initial injury response of HFs, but rather its expression is triggered by that response, whereupon it plays an important role in regeneration. This view is consistent with a recent study showing that various HF regions express chemokines including CCL2, CCL20, and CCL8 in response to stress (Nagao et al., 2012a). The percent HF area with CCL2 expression was highest 1 day after plucking and decreased thereafter (Figure 4E). Unplucked follicles located within (x) or outside (y) of the plucked field showed

low and no CCL2 levels. CCL2 null mice did not express CCL2 after plucking (z).

and accumulates predominantly in plucked follicles (where apoptosis occurs), and to a much smaller extent in neighboring unplucked follicles, which do not undergo apoptosis (Figures 4A, 4B, and S4). This induction is transient and diminishes at day 5. CCL2 induction in plucked HFs occurred regardless of plucking density, so CCL2 was expressed even at the densities that failed to launch regeneration. Epidermal staining of CCL2 is evident in the 2.4 mm specimen, some is seen surrounding the plucked follicle in the 5 mm specimen, but staining is sparse in the 8 mm specimen.

These results are consistent with CCL2 expression providing an overall measure of the extent of plucking and therefore potentially serving as a quorum signal. To test whether CCL2 function is required for follicle regeneration, we waxed whole back skin from both wild-type C57BL/6 and CCL2 null mice. In contrast to the localized plucking of 200 hairs (that induces hair regeneration after ~12 days) (Plikus et al., 2008), wax-stripping the whole back skin drives telogen hairs back into full anagen (anagen VI) within ~6 days (Müller-Röver et al., 2001). However,

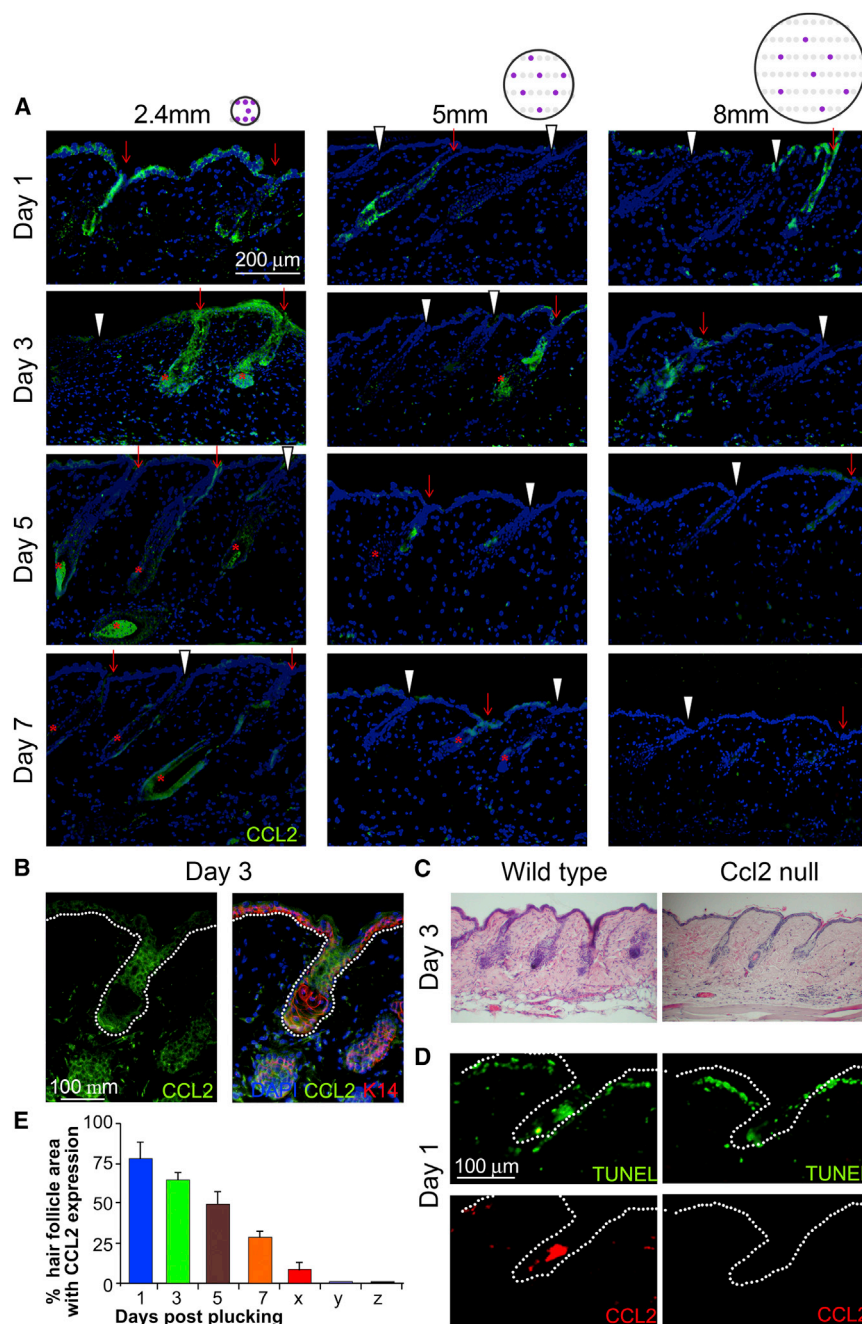
low and no CCL2 levels. CCL2 null mice did not express CCL2 after plucking (z).

### M1 Macrophages Are Mediators Recruited by CCL2 to Execute Quorum-Sensing Behavior

The decay lengths of most diffusible signaling molecules, including chemokines (Sarris et al., 2012; Weber et al., 2013), are much shorter than the decay length we measured for the plucking-induced quorum signal (Figure 2). Chemokines, however, are known to act as chemo-attractants for immune cells, and we postulated that this might play a role in boosting the effective range of action of an initial quorum signal. CCL2 in particular is a potent recruiter of monocyte/macrophage lineage cells.

Indeed, 2 days after plucking, macrophages had heavily infiltrated the plucked skin (Figure 5A). We quantified the macrophage distribution at different times after plucking (Figure 5E). At day 1, F4/80 positive macrophages accumulate around and between the plucked follicles. At day 3, more macrophages spread to the inter-plucked follicular regions and their density





**Figure 4. CCL2 Is Involved in Plucking Induced Hair Regeneration**

(A) HF keratinocytes showed higher CCL2 expression (green) in plucked follicles (red arrow) than in unplucked follicles (white arrowhead). The circle with purple dots indicates the topology of plucked follicles (see also Figure 1E). Peak expression occurs 1–3 days after plucking, and no marked difference between the 2.4, 5, and 8 mm groups were noted. Asterisk represents regenerating HFs.

(B) Double immunostaining for K14 and CCL2 of samples 3 days after plucking showed that HF keratinocytes in plucked follicles are the main source of CCL2.

(C) Hair re-growth is retarded when hairs were plucked from CCL2 null mice.

(D) CCL2 null mice showed similar apoptotic HF cells following plucking as wild-type mice, but could not induce CCL2 in apoptotic HF cells.

(E) Graph showing the percentage of HF area expressing CCL2 at 1, 3, 5, and 7 days post-plucking as well as unplucked HFs within (x) and outside (y) of the plucked field. CCL2 null mice do not express CCL2 (z). (n = 3). Data are represented as mean  $\pm$  SD.

See also Figure S5.

hair plucking. Myeloid cells are nearly absent in normal mice, but are induced at days 3–5 and diminish at day 7 in the transgenic mice (Figure 5F).

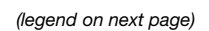
To evaluate their role, we generated a triple transgenic mouse model where myeloid cells are specifically depleted by diphtheria toxin upon doxycycline treatment (*LysM-Cre; Rosa-rtTA; TetO-DTA*). We plucked 200 hairs/5 mm diameter region, which usually launches a quorum sensing response, leading to regeneration. In this mutant, hair regeneration did not occur (Figures 5G and S2D). These myeloid cells represent mainly macrophages, although technically we cannot rule out other cell types completely. All together, the data suggest macrophages play a major role in this process.

Macrophages can be divided into two major types; M1 macrophages (classically activated) exert proinflammatory activities, and M2 macrophages (alternatively activated) are involved in resolving inflammation (Gordon, 2003; Willenborg et al., 2012). Immunostaining showed that M1, but not M2 macrophages, were present 5 days post-plucking (Figures 5B, 5C, and S6A).

These findings are consistent with other studies implicating chemokines in the recruitment of inflammatory macrophages during wound healing and a role for such cells in tissue repair (Willenborg et al., 2012). Since M1 macrophages express CCR4 (Figures 5D and S6A), the receptor for CCL2, we think that these macrophages are recruited to plucked

is substantially elevated (at least four times over background) up to 66% of that found for plucked follicles. The spread can span a distance of 1 mm. These macrophages start to dissipate at day 7 post-plucking.

To test whether macrophages play a functional role in plucking-induced hair regeneration, we used chemical inhibitor and genetic deletion assays. The application of Clodronate liposomes to suppress macrophage function caused an  $\sim$ 12-day delay in hair plucking induced regeneration (Figure 6F). For the genetic approach, we used *LysM-Cre;R26R* transgenic mice to examine the distribution of LacZ positive myeloid cells following



follicles by plucking-induced CCL2. Consistent with this view, plucking failed to induce the accumulation of M1 macrophages in CCL2 null mouse skin (Figure 5H). These data support the model that CCL2 expressed by plucked follicles recruits CCR4-expressing M1 macrophages, which play an essential role in regeneration. We next explored how this comes about.

### Hair Regeneration Induced by Quorum Sensing Is *Tnf- $\alpha$* Dependent

Macrophages are known to produce *Tnf- $\alpha$* . *Tnf- $\alpha$*  mRNA was induced ~2 days after plucking, before anagen initiates (Figures 3B, 3C, and S3). Low power whole mount in situ hybridization reveals that, under conditions in which regeneration occurs, *Tnf- $\alpha$*  is enriched in the extra-follicular environment of both plucked and unplucked hairs (Figure 6A). Double immunostaining showed that these cells are indeed M1 macrophages (Figures 5B, 5C, 5D, 6B, and S6).

Semiquantitative analysis of immunostained specimens obtained from regions of different plucking densities support the view that *Tnf- $\alpha$*  expressing macrophages accumulate around HF's that regenerate after plucking (both plucked and unplucked), but do not accumulate under plucking conditions that fail to activate hair regeneration (Figure S6D).

Quantitative measurements of macrophage-derived *Tnf- $\alpha$*  immunoreactivity over time in a threshold-plucking density region (200 hairs/5 mm diameter) showed that *Tnf- $\alpha$*  positive cells are induced around plucked follicles. They then increased significantly at day 3 and 5 after plucking, spreading into the dermal region between plucked follicles. They decreased at day 7 to approach basal levels (Figures 5E, 6C, and S7B).

To investigate the functional importance of *Tnf- $\alpha$*  in plucking-induced hair regeneration, beads coated with *Tnf- $\alpha$* -related peptide were injected into refractory telogen stage mouse skin. Hair regeneration was induced, followed by propagation to the surrounding region (Figure 6E). Control bead injection did not induce hair regeneration even after 30 days (Figure 6G). Conversely, when hairs were plucked at high density in *Tnf- $\alpha$*  null mice (Figure 6F), a 15-day delay in regeneration was observed. These results indicate that *Tnf- $\alpha$*  is one of the major players for plucking-induced hair regeneration.

Last, we searched for molecules and signals that might function further downstream in hair regeneration. For example, *Tnf- $\alpha$*  is known to stimulate both JNK and NF- $\kappa$ B (nuclear factor

kappa-light-chain-enhancer of activated B cell) signaling. It is also known that activation of the FGF signaling pathway can trigger hair regeneration (Greco et al., 2009). We therefore screened inhibitors of NF- $\kappa$ B, JNK, PI3K, FGF receptor, p38 MAPK, and Erk for effects on plucking-induced hair regeneration. Only NF- $\kappa$ B inhibitors delayed hair regeneration, doing so by 10 days (Figures 6H and S7C). In addition, *Tnf- $\alpha$* -related peptide significantly stimulates the expression of *Wnt3*, *Wnt10a*, and *Wnt10b* in keratinocytes (Figure 6I). Although the *Eda*-NF- $\kappa$ B pathway is important in hair development, a previous study indicated that *Eda* participates in anagen to catagen transition during the postnatal hair regeneration cycle (Fessing et al., 2006). Hence, it is not likely that *Eda* is involved in the plucking induced hair regeneration response. Together the results raise the possibility that *Tnf- $\alpha$* , acting through the NF- $\kappa$ B pathway, ultimately stimulates hair regeneration through activation of Wnt signaling.

## DISCUSSION

### Social Behaviors in an Organ Population

Many organs are composed of repeated, semi-autonomous tissue units, such as acini, crypts, and follicles. The potential for dynamic coupling between the behaviors of such units creates opportunities for collective phenomena. A dramatic example of this is the "hair wave," a coordinated hair cycle wave that can travel across the skin of mammals (Suzuki et al., 2003; Plikus et al., 2008, 2011; Murray et al., 2012).

In this work, we characterize another collective behavior of HF's; density- and topology-dependent, plucking-induced regeneration, which can be viewed as a form of quorum sensing. Quorum sensing is a process whereby a population makes a collective decision based on the number or density of individuals that meet a certain criterion. Typically, a response occurs only when a threshold is exceeded. Quorum sensing has been invoked to describe bacterial cell-to-cell communication (Bassler, 2002) that serves to influence gene regulation in response to population density fluctuations (Miller and Bassler, 2001). Synthetic quorum sensing circuits in yeast were used to demonstrate the diversity of social behaviors that can come from collective communication (Youk and Lim, 2014). Quorum sensing also has been used to explain the collective decision-making behavior of social insects such as ants and honey bees (Pratt, 2005; Visscher, 2007).

### Figure 5. CCL2 Stimulates *Tnf- $\alpha$* Production by Attracting CCR4 (+) M1 Macrophages

- (A) *Tnf- $\alpha$*  is upregulated in the dermal macro-environment on day 2 after wax stripping. *Tnf- $\alpha$*  in the dermal macro-environment is produced by both dermal macrophages (F4/80+ cells, yellow arrow) and adipose cells (red arrow; see also Figure S5C). Few macrophages (yellow arrow) are present at hour 4 and day 10 after plucking. These macrophages do not express *Tnf- $\alpha$* .  
 (B and C) Staining shows that *Tnf- $\alpha$*  is mainly produced by M1 (iNOS-positive) rather than M2 (Arginase-positive) macrophages.  
 (D) *Tnf- $\alpha$*  (+) cells express CCR4 in response to CCL2.  
 (E) The number of F4/80+ cells is highest near plucked follicles and their density decreases with increasing distance from the plucked follicles. See Figure S7A for the unit area we quantified for each data point. The number of F4/80+ cells is rapidly elevated at day 1 post-plucking, reaching a maximum at day 3 and then diminishing at 5 and 7 days after plucking.  
 (F) *LysM-Cre;R26R* reporter mice show that the myeloid lineage-derived cells mostly are induced in the dermis around plucked HF's.  
 (G) When 200 hairs were plucked from myeloid cell-deficient mice from 5 mm region, hairs cannot be induced.  
 (H) *Tnf- $\alpha$*  (+) cells was not induced in CCL2 null mice.  
 (I) *Tnf- $\alpha$*  serum levels are similar between wild-type and CCL2 null mice. Data are represented as mean  $\pm$  SD.  
 See also Figures S6 and S7.



### Molecular Nature of the Quorum-Sensing Circuit

Briefly, the quorum sensing circuit we describe here provides a way for injured HFs to collectively assess the magnitude and extent of injury that the skin has sustained and make an all-or-none decision whether or not to regenerate. A striking feature of this circuit, revealed through molecular modeling, is that the information being shared among follicles decays with a characteristic length of  $\sim 1$  mm, substantially greater than the measured decay lengths of diffusible signaling molecules (Teleman and Cohen, 2000; Müller et al., 2012; Sarris et al., 2012; Weber et al., 2013; Shimozono et al., 2013). The explanation for this apparent paradox seems to reside in the multi-stage nature of the quorum signal, which begins with diffusible molecules, but eventually involves the recruitment of motile cells (inflammatory macrophages) that spread within the tissue. Below we summarize the sequence of molecular and cellular events revealed by the present study (Figure 7).

- (1) Micro-injury and inflammation. Hair plucking leads to hair keratinocyte apoptosis (Ito et al., 2002) (Figure 3A). This in turn leads to inflammatory changes and to the localized overexpression of several inflammatory cytokines, especially CCL2, which may be detected within 12 hr post-plucking (Figures 3B and 4).
- (2) Molecular signal release and dissemination. CCL2 and other cytokines are secreted from plucked follicles and may also involve epidermis around the plucked follicle. The importance of CCL2 is demonstrated by the fact that, in CCL2 null skin, regeneration is markedly delayed (Figure 4). The fact that it is not prevented entirely suggests that some other cytokines induced by plucking may act redundantly with CCL2.
- (3) Recruitment of macrophages as motile vectors. The local production of CCL2 appears to recruit CCR4 (+) M1 macrophages in the dermis (Figure 5). Whereas macrophages initially appear to be enriched around plucked follicles, recruited macrophages soon spread throughout the whole region. By relaying a signaling response with a motile cellular vector, HFs effectively solve the problem of spreading quorum information over long distances.

Another motile vector candidate is the epidermal dendritic Langerhans cell since it also expresses F4/80 antigen. However, F4/80 positive cells appear in dermis at day 1, but do not appear in the epidermis until days 5–7. Our microarray data also did not reveal upregulation of Langerhans cell markers, such as CD207 (Langerin) and CD11b. Although we do not completely rule out the involvement of Langerhans in this process, our data so far suggest a major role for dermal macrophages in this process.

It is worthwhile to mention here that more examples of extended cellular process that mediate signal communication are being identified. For example, in zebrafish stripe pattern formation, pigment cells can utilize their contact-dependent depolarization and repulsive behavior as non-diffusible inhibitors that follow Turing principles (Inaba et al., 2012). In *Drosophila* epithelia, cytonemes can establish a dynamic hedgehog morphogen gradient that may reach afar (Bischoff et al., 2013).

Future in vivo imaging studies of the mouse skin model studied here will allow us to elucidate the interactive cellular behaviors between HFs, immune system and regeneration.

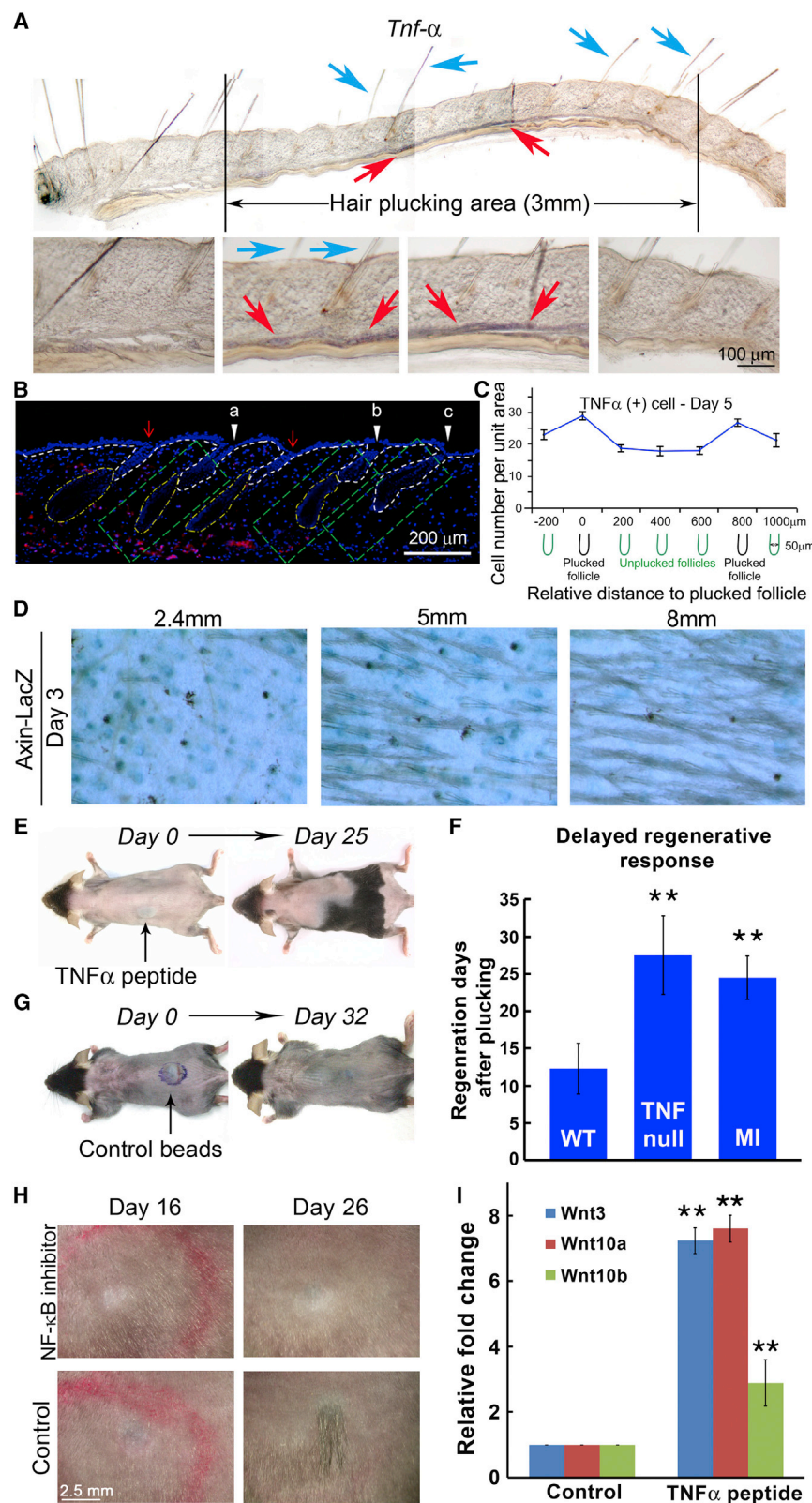
- (4) Release of Tnf- $\alpha$  and collective regeneration. Inflammatory macrophages that are recruited to wound fields secrete Tnf- $\alpha$ , which has been shown to activate hair cycle regeneration (Figure 6E) (Duheron et al., 2011). Regeneration is greatly impaired in Tnf- $\alpha$  null mice (Figure 6F); moreover, Tnf- $\alpha$  serum levels are normal in CCL2 null mice that show impaired plucking-induced regeneration (Figure 5I). These studies indicate that local, not systemic Tnf- $\alpha$  is required for regeneration. Although the exact mechanism by which Tnf- $\alpha$  triggers follicle regeneration is not clear, the data suggest that Tnf- $\alpha$  may act through NF- $\kappa$ B which in turn activate canonical WNT signaling (Figure 6D) (Cawthorn et al., 2007; Schwitalla et al., 2013). While Tnf- $\alpha$  immunoreactivity is mainly in macrophages, it is also detected in other cell types and may provide additional possible mechanisms. The EDAR pathway may also activate NF- $\kappa$ B. However, EDAR it is more involved in anagen/catagen transition (Fessing et al., 2006), not telogen/anagen transition. Further, it is not induced in our microarray data (not shown).

### Adaptive Role of Quorum Sensing

Depending on the severity of skin injury, the body may use different mechanisms to alert, defend, and regenerate the damaged tissue. Plucking of a single hair follicle is a micro-injury. An open, full-thickness wound is a catastrophic event (i.e., macro-injury). While wounded skin is known to induce hair regeneration, small and large wounds may share a fundamental mechanism, but use different molecular circuits to achieve different levels of restoration and regeneration. Indeed, HF activities have been linked to the wound-healing and regenerative behaviors of the inter-follicular epidermis. This link may be mediated by the immune system (Paus et al., 1998), macrophage recruitment (Osaka et al., 2007), and increased Tnf- $\alpha$  expression (Jiang et al., 2010). Interestingly, TNF- $\alpha$  converting enzyme, a regulator of Tnf- $\alpha$ , is a component of the HF bulge niche (Nagao et al., 2012b). Anagen phase HFs can influence the surrounding epidermis to markedly accelerate wound healing (Ansell et al., 2011). In mice, loss of full thickness skin larger than 1 cm in diameter could lead to new follicle formation (Ito et al., 2007). However, plucking does not launch a full wound healing response, so conceptually plucking works differently from the wound and our study focuses at a different scale. It provides a novel understanding into how HFs respond to injury at the level of a HF population. We analyzed how interactions among HFs and the dermal environment reach a binary choice based on a collective measurement of injury. We show that effective damage control is achieved via co-option of existing signaling mechanisms (e.g., Tnf- $\alpha$ , macrophage) for the “social behaviors” of a stem cell population.

In summary, we report a higher level integration of signals from hair regeneration, immune cytokines, and wound healing. Instead of a top-down process, quorum sensing represents a bottom-up process based on local information. Each follicle





**Figure 6. Hair Regeneration Is Proportional to the Local Concentration of *Tnf-α***

(A) Whole mount in situ hybridization shows *Tnf-α* (brown color in the dermis, red arrows) was induced under plucked and unplucked hairs (blue arrows) toward the center of the 3 mm plucked zone 5 days after plucking.

(B) Semiquantitative assessment of the *Tnf-α* concentration using the 5mm group at day 5. Its expression level was quantified in three different skin regions (green boxes) that differ in their proximity to plucked follicles. "a" is closest to the plucked follicles and shows the highest *Tnf-α* levels. "b" is away from the plucked follicles and shows lower *Tnf-α* levels. "c" is furthest away and shows the least *Tnf-α*.

(C) Quantitative assessment of the *Tnf-α* positive cells around plucked follicles and inter-plucked follicle dermis. See Figure S7B for complete series. The pattern is similar to that of F4/80 macrophage distribution ( $n = 3$ ).

(D) Density-dependent plucking on *Axin-LacZ* mice show that the canonical Wnt/ $\beta$ -catenin signaling pathway was activated 3 days after plucking and the number of LacZ (+) HF s was proportional to the plucking density.

(E) Subcutaneous injection of *Tnf-α*-related peptide coated beads during refractory telogen can induce anagen re-entry and then propagate to the surrounding HF s.

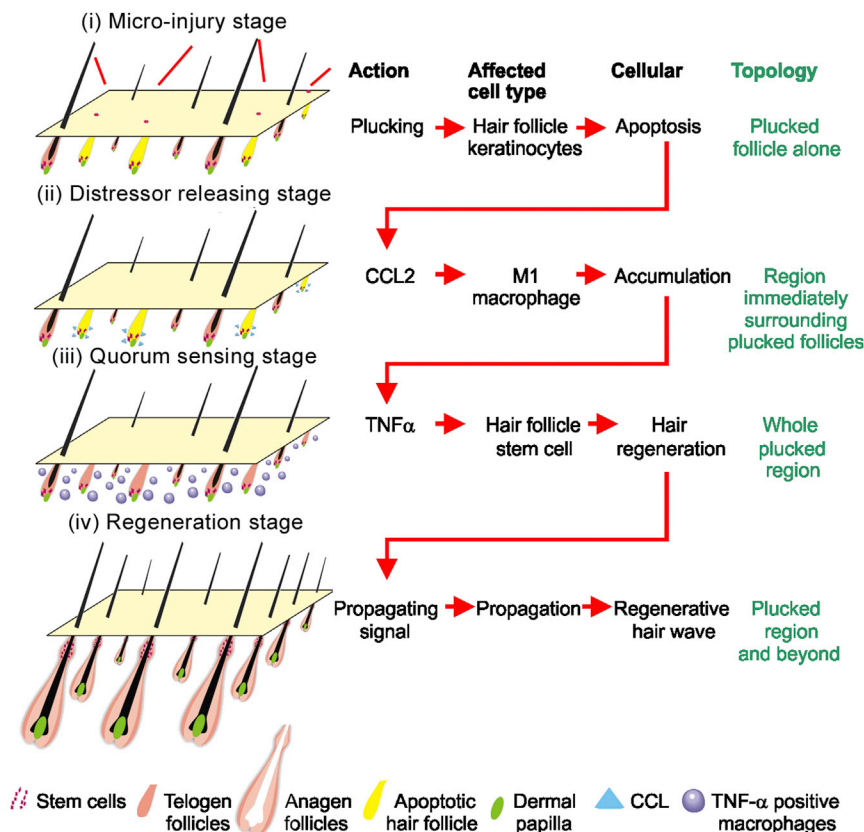
(F) *Tnf-α* null mice exhibit a 15-day delay in anagen re-entry following plucking of 200 hairs during refractory telogen phase. Intra-peritoneal macrophage inhibitor (MI) injection can also delay plucking induced hair regeneration by 12 days.

(G) Albumin coated beads injection showed no anagen re-entry even after 32 days.

(H) Subcutaneous NF- $\kappa$ B inhibitor injection can delay plucking-induced hair regeneration by 10 days.

(I) Wnt3, Wnt10a, and Wnt10b were activated in keratinocytes by TNF-related peptides. Data are represented as mean  $\pm$  SD. \*\* $p < 0.001$ .

See also Figures S6 and S7.



**Figure 7. Molecular Basis of Quorum-Sensing Behavior during the Activation of Hair Stem Cells in the Follicle Population**

Schematic illustration of the process. Stage i: minor injury → hair keratinocyte apoptosis → CCL2 production. Stage ii: CCL2 secretion → macrophage accumulation. Stage iii: macrophage and  $Tnf-\alpha$  permeate the whole region. Stage iv:  $Tnf-\alpha$  activates hair regeneration in the whole region. Hair regeneration further spreads due to propagation of regenerative hair waves. Please see text for more detail of the model.

becomes a sensor for the population to assess the level of damage. The molecular circuit quantifies injury strength by summing together local signals from different organs. Here, the communication among tissues reaches a larger scale organization by coupling local molecular signaling (in the form of a chemical gradient) with motile cellular vectors. In this study, macrophages are identified as a motile vector that allows a length scale of up to 1 mm. In this manner, the injury response is measured and reflects local needs. This study may just be one of the examples that reveal collective cellular behaviors in response to physiological or pathological stimuli. We believe that the quorum sensing behavior principle is likely to be present in the regeneration of tissue and organs beyond the skin.

## EXPERIMENTAL PROCEDURES

### Surgical Procedures

All procedures were performed on anesthetized animals with protocols approved by the University of Southern California Institutional Animal Care and Use Committee (USC IACUC). Hair cycle was synchronized by wax stripping (Müller-Röver et al., 2001). Hairs in refractory telogen were plucked with the spacing indicated in the result section. Regenerative hair numbers are counted under a dissection microscope.

### RNA Preparation and Microarray

For microarray, all the dermal tissues are collected. RNA was prepared using TRI Reagent BD (Sigma-Aldrich) following the manufacturer's recommendations. Please see Extended Experimental Procedures for detail.

The microarray data reported here have been submitted to the GEO (accession number GSE46181). Primer sequences for RT-PCR are listed in Table S1.

### Perturbation of Quantitative Plucking

Small molecular inhibitor or peptides were injected intra-dermally on one side of mouse dorsal skin for 4 days. Then 200 hairs were plucked in the center of the injected area. After plucking, these drugs were continuously injected for an additional 6 days. DMEM was injected to the opposite side as a control. Each animal was injected with only one reagent.

### ACCESSION NUMBERS

The GEO accession number for the microarray data reported in this paper is GSE46181.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.016>.

### AUTHOR CONTRIBUTIONS

C.C.C. and C.M.C. conceived the overall experimental design. C.C.C. and T.X.J. did the hair plucking and characterization. L.W., S.S., and M.W.H. did immune characterization. A.D.L. and P.J.M. did mathematical modeling. A.D.L. did mathematical estimation of length scale for quorum sensing. M.V.P., R.R., and C.F.G.-J. did LysM-Cre mouse work. C.C.C., C.M.C., A.D.L., M.V.P., R.B.W., and O.L. did manuscript writing and editing.

## ACKNOWLEDGMENTS

C.M.C., T.X.J., and R.B.W. are supported by National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) R01-AR42177, AR 47364, and AR60306. C.C.C. is supported by NSC 100-2314-B-075-044, NSC 101-2314-B-075-008-MY3, and the Taipei Veterans General Hospital (VN103-12, V103C-010, V102B-009, R-1100403). A.D.L. is supported by NIGMS P50-GM076516. S.S. is supported by NIH R01DE17449. L.W. is supported by the National Natural Science Foundation of China (81270015). M.V.P. is supported by NIAMS R01-AR067273 and Edward Mallinckrodt Jr. Foundation grant. R.R. is supported by a California Institute of Regenerative Medicine (CIRM) training grant (TG2-01152). C.F.G.-J. is supported by the National Science Foundation Graduate Research Fellowship Program (DGE-1321846). P.J.M. is supported by a post-doctoral and early career researcher exchange fellowship from the Northern Research Partnership (NRP). M.W.H. is supported by Top Notch University plan of CKU, Taiwan. We thank Drs. Jaw-Ching Wu, Han-Nan Liu, and Yun-Ting Chang of National Yang Ming University/Taipei Veterans General hospital for their support. Invention number 2014-255 "Enhance hair growth via plucking" (D2014-0054) was disclosed to University of Southern California.

Received: May 22, 2014

Revised: January 12, 2015

Accepted: February 2, 2015

Published: April 9, 2015

## REFERENCES

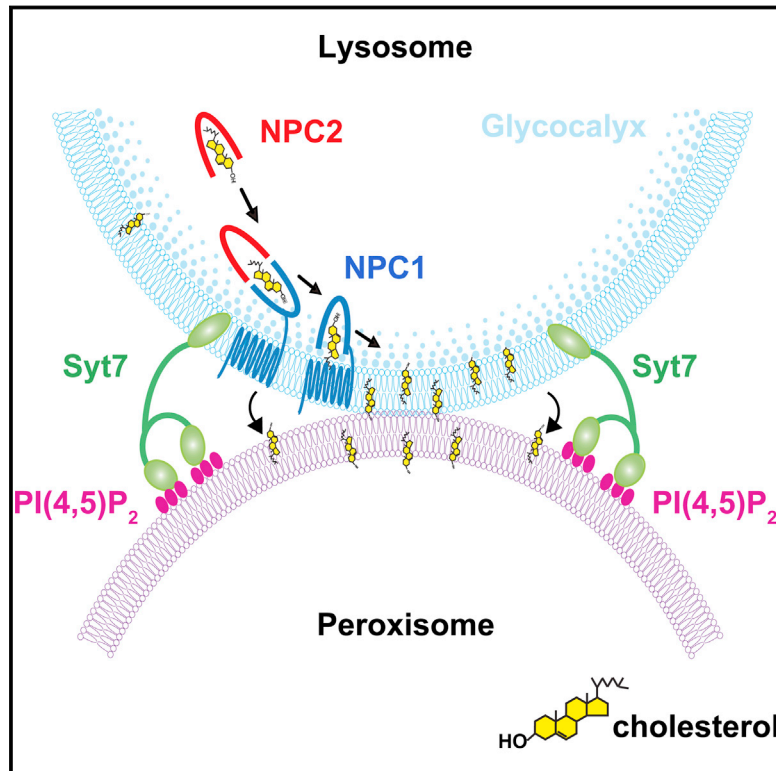
- Ansell, D.M., Kloepper, J.E., Thomason, H.A., Paus, R., and Hardman, M.J. (2011). Exploring the "hair growth-wound healing connection": anagen phase promotes wound re-epithelialization. *J. Invest. Dermatol.* **131**, 518–528.
- Bassler, B.L. (2002). Small talk. Cell-to-cell communication in bacteria. *Cell* **109**, 421–424.
- Bischoff, M., Gradilla, A.C., Seijo, I., Andrés, G., Rodríguez-Navas, C., González-Méndez, L., and Guerrero, I. (2013). Cytonemes are required for the establishment of a normal Hedgehog morphogen gradient in *Drosophila* epithelia. *Nat. Cell Biol.* **15**, 1269–1281.
- Cawthorn, W.P., Heyd, F., Hegyi, K., and Sethi, J.K. (2007). Tumour necrosis factor- $\alpha$  inhibits adipogenesis via a beta-catenin/TCF4(TCF7L2)-dependent pathway. *Cell Death Differ.* **14**, 1361–1373.
- Chen, C.C., and Chuong, C.M. (2012). Multi-layered environmental regulation on the homeostasis of stem cells: the saga of hair growth and alopecia. *J. Dermatol. Sci.* **66**, 3–11.
- Chuong, C.M., Randall, V.A., Widelitz, R.B., Wu, P., and Jiang, T.X. (2012). Physiological regeneration of skin appendages and implications for regenerative medicine. *Physiology (Bethesda)* **27**, 61–72.
- Collins, H.H. (1918). Studies of normal moult and of artificially induced regeneration of pelage in *Peromyscus*. *J. Exp. Zool.* **27**, 73–99.
- Duheron, V., Hess, E., Duval, M., Decossas, M., Castaneda, B., Klöpper, J.E., Amoasii, L., Barbaroux, J.B., Williams, I.R., Yagita, H., et al. (2011). Receptor activator of NF-kappaB (RANK) stimulates the proliferation of epithelial cells of the epidermo-pilosebaceous unit. *Proc. Natl. Acad. Sci. USA* **108**, 5342–5347.
- Enshell-Seijffers, D., Lindon, C., Kashiwagi, M., and Morgan, B.A. (2010). beta-catenin activity in the dermal papilla regulates morphogenesis and regeneration of hair. *Dev. Cell* **18**, 633–642.
- Fessing, M.Y., Sharova, T.Y., Sharov, A.A., Atoyan, R., and Botchkarev, V.A. (2006). Involvement of the Edar signaling in the control of hair follicle involution (catagen). *Am. J. Pathol.* **169**, 2075–2084.
- Festa, E., Fretz, J., Berry, R., Schmidt, B., Rodeheffer, M., Horowitz, M., and Horsley, V. (2011). Adipocyte lineage cells contribute to the skin stem cell niche to drive hair cycling. *Cell* **146**, 761–771.
- Gordon, S. (2003). Alternative activation of macrophages. *Nat. Rev. Immunol.* **3**, 23–35.
- Greco, V., Chen, T., Rendl, M., Schober, M., Pasolli, H.A., Stokes, N., Dela Cruz-Racelis, J., and Fuchs, E. (2009). A two-step mechanism for stem cell activation during hair regeneration. *Cell Stem Cell* **4**, 155–169.
- Inaba, M., Yamanaka, H., and Kondo, S. (2012). Pigment pattern formation by contact-dependent depolarization. *Science* **335**, 677.
- Ito, M., Kizawa, K., Toyoda, M., and Morohashi, M. (2002). Label-retaining cells in the bulge region are directed to cell death after plucking, followed by healing from the surviving hair germ. *J. Invest. Dermatol.* **119**, 1310–1316.
- Ito, M., Yang, Z., Andl, T., Cui, C., Kim, N., Millar, S.E., and Cotsarelis, G. (2007). Wnt-dependent de novo hair follicle regeneration in adult mouse skin after wounding. *Nature* **447**, 316–320.
- Jahoda, C.A., and Christiano, A.M. (2011). Niche crosstalk: intercellular signals at the hair follicle. *Cell* **146**, 678–681.
- Jiang, S., Zhao, L., Teklemariam, T., and Hantash, B.M. (2010). Small cutaneous wounds induce telogen to anagen transition of murine hair follicle stem cells. *J. Dermatol. Sci.* **60**, 143–150.
- Lander, A.D. (2007). Morpheus unbound: reimagining the morphogen gradient. *Cell* **128**, 245–256.
- Lowry, W.E., Blanpain, C., Nowak, J.A., Guasch, G., Lewis, L., and Fuchs, E. (2005). Defining the impact of beta-catenin/Tcf transactivation on epithelial stem cells. *Genes Dev.* **19**, 1596–1611.
- Miller, M.B., and Bassler, B.L. (2001). Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**, 165–199.
- Müller, P., Rogers, K.W., Jordan, B.M., Lee, J.S., Robson, D., Ramanathan, S., and Schier, A.F. (2012). Differential diffusivity of Nodal and Lefty underlies a reaction-diffusion patterning system. *Science* **336**, 721–724.
- Müller-Röver, S., Handjiski, B., van der Veen, C., Eichmüller, S., Foitzik, K., McKay, I.A., Stenn, K.S., and Paus, R. (2001). A comprehensive guide for the accurate classification of murine hair follicles in distinct hair cycle stages. *J. Invest. Dermatol.* **117**, 3–15.
- Murray, P.J., Plikus, M.V., Maini, P.K., Choung, C.M., and Baker, R.E. (2012). Modelling hair follicle growth dynamics as an excitable medium. *PLoS Comp. Biol.* **8**, e1002804.
- Nagao, K., Kobayashi, T., Moro, K., Ohyama, M., Adachi, T., Kitashima, D.Y., Ueha, S., Horiuchi, K., Tanizaki, H., Kabashima, K., et al. (2012a). Stress-induced production of chemokines by hair follicles regulates the trafficking of dendritic cells in skin. *Nat. Immunol.* **13**, 744–752.
- Nagao, K., Kobayashi, T., Ohyama, M., Akiyama, H., Horiuchi, K., and Amagai, M. (2012b). Brief report: requirement of TACE/ADAM17 for hair follicle bulge niche establishment. *Stem Cells* **30**, 1781–1785.
- Osaka, N., Takahashi, T., Murakami, S., Matsuzawa, A., Noguchi, T., Fujiwara, T., Aburatani, H., Moriyama, K., Takeda, K., and Ichijo, H. (2007). ASK1-dependent recruitment and activation of macrophages induce hair growth in skin wounds. *J. Cell Biol.* **176**, 903–909.
- Paus, R., van der Veen, C., Eichmüller, S., Kopp, T., Hagen, E., Müller-Röver, S., and Hofmann, U. (1998). Generation and cyclic remodeling of the hair follicle immune system in mice. *J. Invest. Dermatol.* **111**, 7–18.
- Plikus, M.V., Mayer, J.A., de la Cruz, D., Baker, R.E., Maini, P.K., Maxson, R., and Chuong, C.M. (2008). Cyclic dermal BMP signalling regulates stem cell activation during hair regeneration. *Nature* **451**, 340–344.
- Plikus, M.V., Baker, R.E., Chen, C.C., Fare, C., de la Cruz, D., Andl, T., Maini, P.K., Millar, S.E., Widelitz, R., and Chuong, C.M. (2011). Self-organizing and stochastic behaviors during the regeneration of hair stem cells. *Science* **332**, 586–589.
- Pratt, S.C. (2005). Quorum sensing by encounter rates in the ant *Temnothorax alpeipennis*. *Behav. Ecol.* **16**, 488–496.
- Sarris, M., Masson, J.B., Maurin, D., Van der Aa, L.M., Boudinot, P., Lortat-Jacob, H., and Herbolme, P. (2012). Inflammatory chemokines direct and restrict leukocyte migration within live tissues as glycan-bound gradients. *Curr. Biol.* **22**, 2375–2382.

- Schwitalla, S., Fingerle, A.A., Cammareri, P., Nebelsiek, T., Göktuna, S.I., Ziegler, P.K., Canli, O., Heijmans, J., Huels, D.J., Moreaux, G., et al. (2013). Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. *Cell* **152**, 25–38.
- Shimozono, S., Imura, T., Kitaguchi, T., Higashijima, S., and Miyawaki, A. (2013). Visualization of an endogenous retinoic acid gradient across embryonic development. *Nature* **496**, 363–366.
- Silver, A.F., and Chase, H.B. (1970). DNA synthesis in the adult hair germ during dormancy (telogen) and activation (early anagen). *Dev. Biol.* **21**, 440–451.
- Stenn, K.S., and Paus, R. (2001). Controls of hair follicle cycling. *Physiol. Rev.* **81**, 449–494.
- Suzuki, N., Hirata, M., and Kondo, S. (2003). Traveling stripes on the skin of a mutant mouse. *Proc. Natl. Acad. Sci. USA* **100**, 9680–9685.
- Teleman, A.A., and Cohen, S.M. (2000). DPP gradient formation in the *Drosophila* wing imaginal disc. *Cell* **103**, 971–980.
- Visscher, P.K. (2007). Group decision making in nest-site selection among social insects. *Annu. Rev. Entomol.* **52**, 255–275.
- Weber, M., Hauschild, R., Schwarz, J., Moussion, C., de Vries, I., Legler, D.F., Luther, S.A., Bollenbach, T., and Sixt, M. (2013). Interstitial dendritic cell guidance by haptotactic chemokine gradients. *Science* **339**, 328–332.
- Willenborg, S., Lucas, T., van Loo, G., Knipper, J.A., Krieg, T., Haase, I., Brachvogel, B., Hammerschmidt, M., Nagy, A., Ferrara, N., et al. (2012). CCR2 recruits an inflammatory macrophage subpopulation critical for angiogenesis in tissue repair. *Blood* **120**, 613–625.
- Youk, H., and Lim, W.A. (2014). Secreting and sensing the same molecule allows cells to achieve versatile social behaviors. *Science* **343**, 1242782.



# Cholesterol Transport through Lysosome-Peroxisome Membrane Contacts

## Graphical Abstract



## Authors

Bei-Bei Chu, Ya-Cheng Liao, ...,  
Bo-Liang Li, Bao-Liang Song

## Correspondence

blsong@whu.edu.cn

## In Brief

Lysosome forms dynamic membrane contacts with peroxisome, and cholesterol is transported from lysosome to peroxisome. Massive cholesterol accumulates in the cells from patients with peroxisomal disorders.

## Highlights

- Genome-wide RNAi screen reveals 341 genes important for cholesterol transport
- Lysosomal Syt7 binds peroxisomal PI(4,5)P<sub>2</sub> to bridge the organelle contact
- Organelle contacts mediate cholesterol transport from lysosome to peroxisome
- Cholesterol is accumulated in cells and animal models of peroxisomal disorders



# Cholesterol Transport through Lysosome-Peroxisome Membrane Contacts

Bei-Bei Chu,<sup>1,2,3,5</sup> Ya-Cheng Liao,<sup>1,5</sup> Wei Qi,<sup>1</sup> Chang Xie,<sup>1</sup> Ximing Du,<sup>4</sup> Jiang Wang,<sup>3</sup> Hongyuan Yang,<sup>4</sup> Hong-Hua Miao,<sup>1</sup> Bo-Liang Li,<sup>1</sup> and Bao-Liang Song<sup>2,\*</sup>

<sup>1</sup>State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

<sup>2</sup>College of Life Sciences, the Institute for Advanced Studies, Wuhan University, Wuhan 430072, China

<sup>3</sup>College of Animal Sciences and Veterinary Medicine, Henan Agricultural University, Zhengzhou 450002, Henan Province, China

<sup>4</sup>School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia

<sup>5</sup>Co-first author

\*Correspondence: [blsong@whu.edu.cn](mailto:blsong@whu.edu.cn)

<http://dx.doi.org/10.1016/j.cell.2015.02.019>

## SUMMARY

Cholesterol is dynamically transported among organelles, which is essential for multiple cellular functions. However, the mechanism underlying intracellular cholesterol transport has remained largely unknown. We established an amphotericin B-based assay enabling a genome-wide shRNA screen for delayed LDL-cholesterol transport and identified 341 hits with particular enrichment of peroxisome genes, suggesting a previously unappreciated pathway for cholesterol transport. We show dynamic membrane contacts between peroxisome and lysosome, which are mediated by lysosomal Synaptotagmin VII binding to the lipid PI(4,5)P<sub>2</sub> on peroxisomal membrane. LDL-cholesterol enhances such contacts, and cholesterol is transported from lysosome to peroxisome. Disruption of critical peroxisome genes leads to cholesterol accumulation in lysosome. Together, these findings reveal an unexpected role of peroxisome in intracellular cholesterol transport. We further demonstrate massive cholesterol accumulation in human patient cells and mouse model of peroxisomal disorders, suggesting a contribution of abnormal cholesterol accumulation to these diseases.

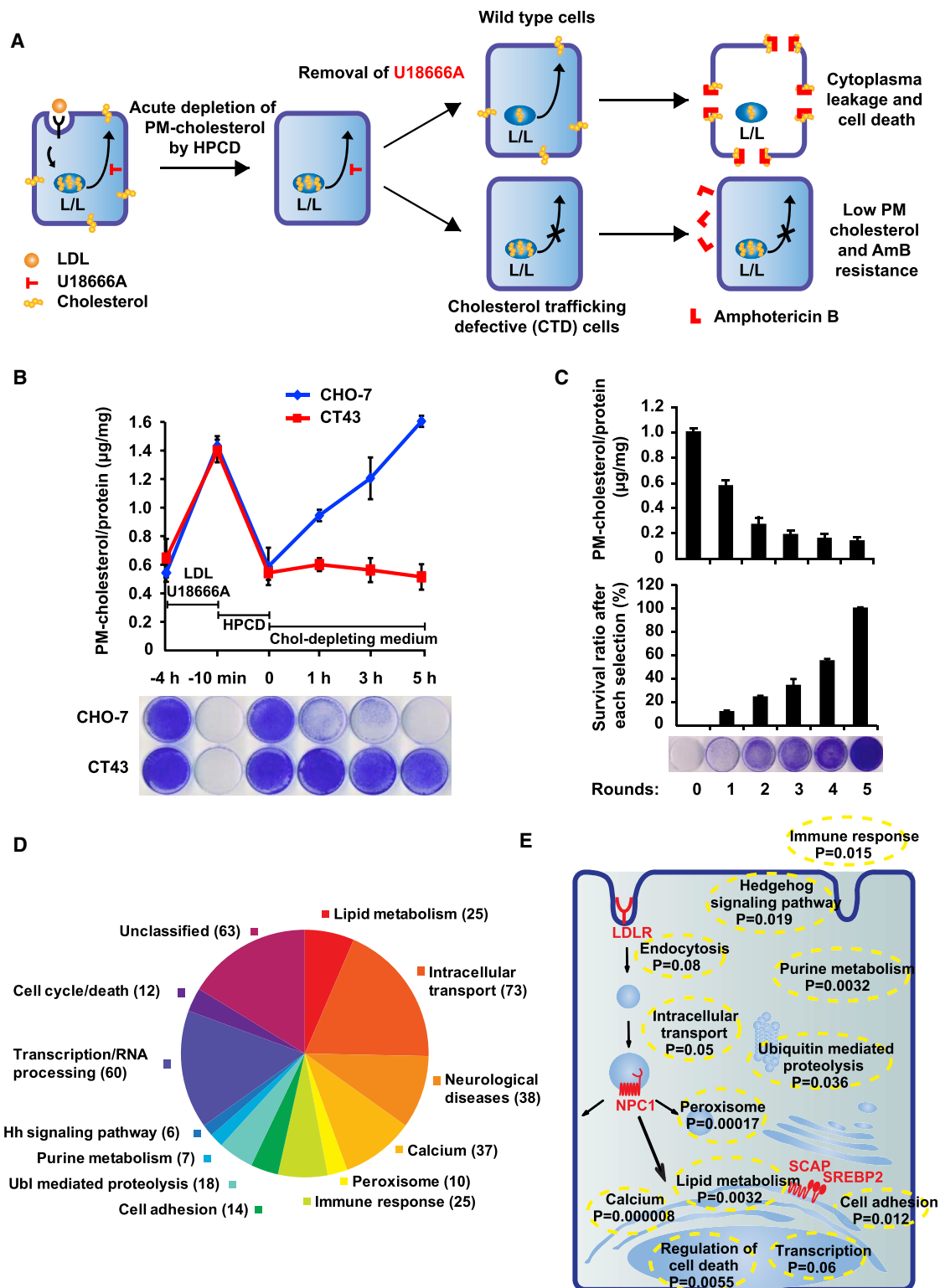
## INTRODUCTION

Cholesterol, an essential lipid for eukaryotic cells, plays important roles in many cellular processes including membrane properties regulation, steroidogenesis, bile acid synthesis, and signal transduction. Accounting for ~30%–40% of total cellular lipids, cholesterol is dynamically transported in cells and unevenly distributed in cellular membrane structures. Only ~0.5%–1% of total cellular cholesterol is present in the ER membrane (Lange et al., 1999) and its concentration is higher in the Golgi apparatus and highest (~60%–80%) in the plasma membrane (PM) (Liscum and Munn, 1999). In addition, cholesterol exerts diverse cellular

functions in different organelles. Sterols in ER control de novo cholesterol biosynthesis by inhibiting SREBP processing and promoting degradation of HMG-CoA reductase (Goldstein et al., 2006). Cholesterol is esterified in ER for storage and lipoprotein secretion (Chang et al., 1997; Vance and Vance, 1990) and oxidized and converted to steroids and bile acids in mitochondria and peroxisome (Ishibashi et al., 1996). Thus, dynamic cholesterol transport in cells is pivotal for multiple cellular functions.

Low density lipoprotein (LDL)-derived cholesterol trafficking is a major part of intracellular cholesterol transport with most mammalian cells acquiring ~80% of their cholesterol through receptor-mediated endocytosis of plasma LDL (Brown and Goldstein, 1986). Upon receptor binding and internalization, LDL is delivered from early endosome to late endosome/lysosome (L/L), where LDL-derived cholesteryl esters are hydrolyzed to unesterified cholesterol. Free cholesterol then egresses from L/L and is further passed to downstream organelles such as the PM, ER, and mitochondria to fulfill its functions (Chang et al., 2006). To date, most mechanistic knowledge on cholesterol passage from L/L to other organelles has come from studies of the inheritable neuronal degeneration disorder Niemann Pick type C (NPC) disease, which is caused by loss-of-function mutations in *NPC1* or *NPC2* genes (Carstea et al., 1997; Sleat et al., 2004). NPC patients show severe cholesterol accumulation in multiple tissues. NPC1 is a polytopic membrane protein on L/L, whereas NPC2 is a luminal protein. After cholesteryl ester is hydrolyzed in the lysosomal lumen, NPC2 binds the unesterified cholesterol by recognizing the 8-carbon isooctyl side chain. NPC2 then hands over the cholesterol molecule to the N-terminal domain of NPC1, with the 3 $\beta$ -hydroxyl group buried within the binding pocket. The NPC1-bound cholesterol projects through the glycolyx and is inserted into the lysosomal membrane. In *NPC1* or *NPC2* mutant cells, cholesterol cannot be incorporated into membrane and is therefore accumulated in the lumen (Kwon et al., 2009). However, this only accounts for how free cholesterol reaches the L/L membrane, and the mechanisms whereby cholesterol leaves the lysosomal membrane and moves to other organelles remain largely unknown.

To identify critical proteins for intracellular cholesterol transport, we developed a cellular system using the antifungal



**Figure 1. Genome-wide RNAi Screen Identifies Genes Involved in Intracellular Cholesterol Transport**

(A) Schematic representation of the screen strategy.

(B) The cells were treated as shown in (A) and Figure S1C. The PM cholesterol and effect of AmB on cell growth at each time point were determined.

(C) PM cholesterol levels and survival ratio based on crystal violet staining of each selection round. Results represent the mean  $\pm$  SD of three independent experiments.

(legend continued on next page)

antibiotic amphotericin B (AmB), in which cells only survive when they have impaired intracellular cholesterol transport. We performed a genome-wide pooled shRNA screen with the AmB system and identified over 300 genes affecting cholesterol transport. The genes encoding peroxisomal proteins were enriched. We further demonstrated that peroxisome forms transient lysosome-peroxisome membrane contact (LPMC) with lysosome through the binding of peroxisomal lipid PI(4,5)P<sub>2</sub> by lysosomal protein Synaptotagmin VII (SyT7). Cholesterol can be transported to peroxisome from lysosome through LPMC. Consistent with the latter findings, we observed drastic cholesterol accumulation in the X-chromosomal form of adrenoleukodystrophy (X-ALD) mouse model and in fibroblasts from human patients with different types of peroxisomal disorders. Our findings therefore reveal a fundamental role of peroxisome in intracellular cholesterol transport and suggest potential novel strategies for the diagnosis and treatment of peroxisome-related diseases.

## RESULTS

### Genome-wide Pooled shRNA Screening for Cholesterol Trafficking Defective Cells

AmB binds to cholesterol in PM and forms pores that lead to cytoplasm leakage and cell death (Andreoli, 1973). Based on this property, we designed a genome-wide shRNA screen to identify genes required for intracellular cholesterol transport, in particular the transport of cholesterol from LDL receptor (LDLR)-mediated endocytosis. The rationale and overall process of the screen are depicted in Figure 1A. There are three key elements, namely: (1) inhibition of endogenous cholesterol biogenesis throughout the entire process and delivery of cholesterol by LDL particles to focus on the transport of LDL-derived cholesterol, (2) synchronization of cells at the stage of high cholesterol in L/L and low cholesterol in PM so that the cholesterol can be transported to the PM in all cells at a given time point, and (3) enrichment of cholesterol trafficking defective (CTD) cells by using AmB that kills the cells with proper cholesterol transport in a controlled manner. The first key element is achieved by using lovastatin to inhibit HMG-CoA reductase and low concentration of mevalonate to only permit the synthesis of nonsterol isoprenoids essential for cell growth. Lipoprotein-deficient serum is also used before LDL delivery so that the cells are in cholesterol starvation and the initial LDLR level is very high. The second key element is realized by using U18666A, a compound that reversibly blocks cholesterol efflux from L/L (Liscum and Faust, 1989), and cyclodextrin, a cholesterol mobilizing reagent (Liu et al., 2010; Rosenbaum et al., 2010). Co-treatment of cholesterol-starved cells with LDL, lovastatin, and U18666A leads to LDLR-mediated endocytosis of large amounts of cholesterol which is trapped in L/L by U18666A. After a short exposure to cyclodextrin to acutely deplete cholesterol from PM, the cells are incubated without U18666A to allow cholesterol transport

from L/L to PM. AmB is then used to kill the cells with more cholesterol in PM. The cholesterol trafficking rate and PM-cholesterol level are lower in CTD cells than wild-type (WT) cells at particular time points. Thus, these CTD cells can survive AmB treatment.

The procedure described above was validated by comparing WT CHO-7- and *NPC1*-deficient CT43 cells (Figure 1B). Cyclodextrin decreased the PM-cholesterol level to 0.59  $\mu\text{g}/\text{mg}$  protein. After removal of U18666A, PM-cholesterol level was much higher in CHO-7 than CT43 cells and the former was more sensitive to AmB treatment (Figure 1B). To perform the screen, HeLa cells were infected with a pooled shRNA library and the virus-infected cells were subjected to AmB selection as described above (Figure S1A). We observed gradual decrease of PM-cholesterol and increase of survival rate in the first five rounds of selection before reaching plateau (Figure 1C), suggesting that CTD cells were largely enriched. The shRNA inserts were then amplified from the CTD cells and subjected to deep sequencing.

The RNAi screening identified 341 candidate genes, each of which was targeted by two or more small hairpin RNAs (shRNAs), eliminating the off-target effect of shRNA. Their symbols and basic information are listed in Table S1.

### Analysis and Validation of Screening Results

To characterize the enriched biological processes and pathways in our screen, the 341 gene hits were subjected to gene ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) database analysis (Figures 1D and 1E). The genes involved in lipid metabolism and intracellular transport were amply presented, constituting 28.7% of total candidates (Figures 1D and 1E). Among these hits, there is *NPC1*, loss of which is well known to trap cholesterol in lysosome and prevent cholesterol from traveling to PM. This serves as a positive control and suggests our screen was successful. Our screen also recovered genes that participate in *LDLR* expression regulation and endocytosis, such as *SREBP2*, *SCAP* (Brown and Goldstein, 1997), *LDLR* (Brown and Goldstein, 1986), and *AP2 associated kinase 1* (*AAK1*) (Conner and Schmid, 2002). Because silencing of these genes prevents cells from taking up LDL, their appearance in the candidates list was expected.

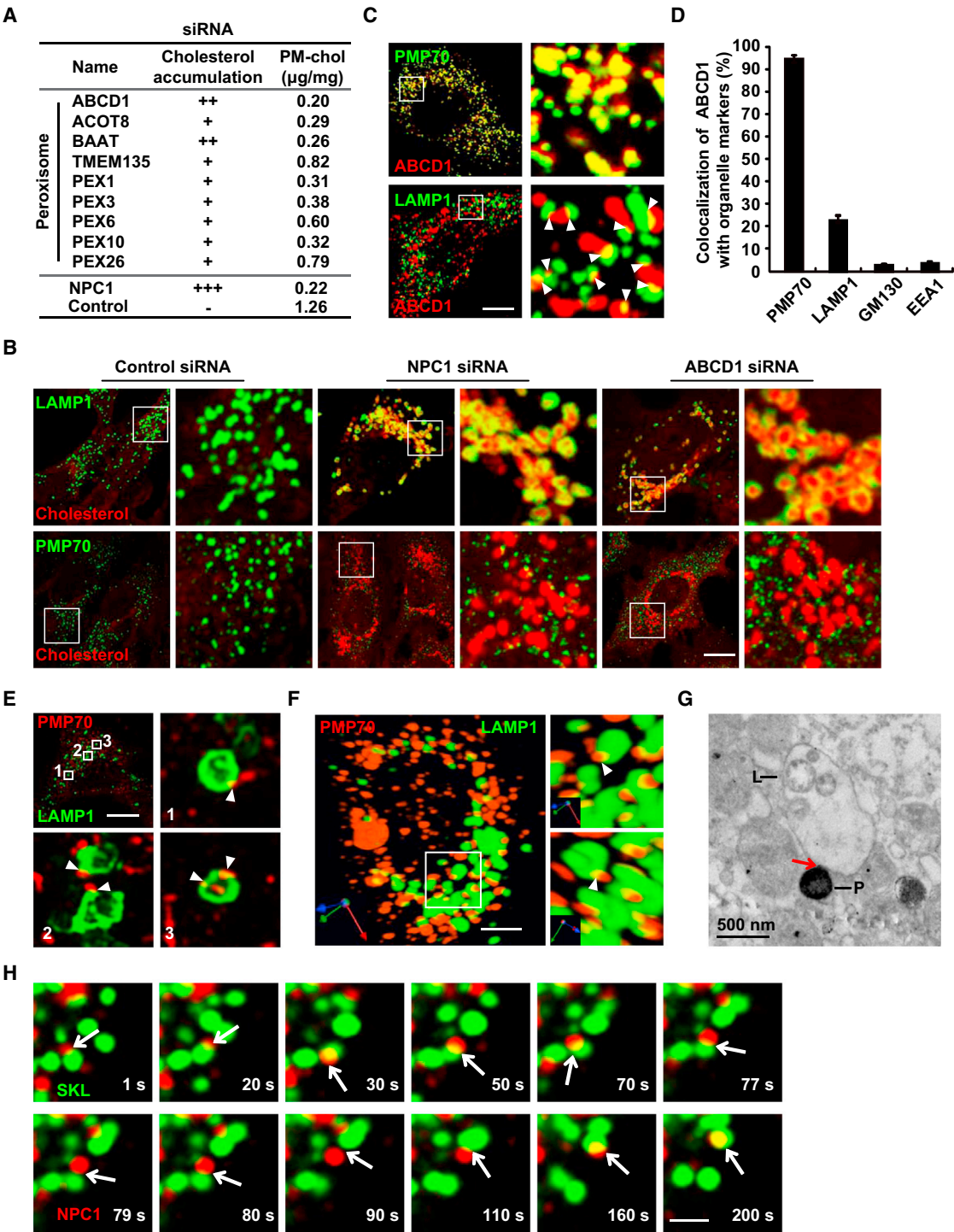
Unexpectedly, we found marked enrichment for genes associated with neurological diseases, peroxisome, calcium, transcription/RNA processing, immune response, cell adhesion, Hh pathway, ubiquitin-mediated proteolysis, and purine metabolism. It is interesting that neurological disease-related genes are discovered in our screen to affect cholesterol transport. As exemplified by NPC disease, which is characterized by severe neurological symptoms secondary to cholesterol accumulation in lysosome, the neuron is particularly sensitive to cholesterol alteration, and impaired cholesterol transport may be a mechanism shared by these neurological diseases.

(D) Bioinformatics classification of the hits into biological processes and molecular functions categories. The number in the bracket shows the number of genes in each category.

(E) Statistically enriched biological processes superimposed on a sketch depicting a cell, with the corresponding p value of GO analysis in the screen. Genes in red refer to representative hits.

See also Figure S1 and Table S1.





**Figure 2. Peroxisome Forms Transient and Dynamic Membrane Contacts with Lysosome**

(A) Knockdown of the peroxisome genes identified in the screen led to cholesterol accumulation and decrease of PM cholesterol levels. The “+” indicates the degree of cholesterol accumulation; the “-” indicates no obvious cholesterol accumulation.

(B) SV589 cells transfected with indicated siRNAs were stained with filipin (red) and antibody against endogenous LAMP1 (green) or PMP70 (green). Scale bar, 10 μm. LAMP1: lysosome marker, PMP70: peroxisome marker.

(C) HeLa cells transfected with mouse ABCD1-mCherry were assessed by immunostaining with antibody against PMP70 (green) or LAMP1 (green). Scale bar, 10 μm.

(D) Quantification of colocalization of ABCD1 with organelle-specific markers shown in (C) and Figure S3A. GM130, Golgi marker; EEA1, early endosome marker. Data represent mean ± SD (n = 4, 35 cells per independent experiment).

(legend continued on next page)

To further confirm the hits, we selected 30 representative genes covering all 14 classes and validated them using distinct shRNA sequences. The survival rate of knockdown cells is dramatically higher upon AmB treatment as compared with control cells (Figure S1D). Among the 30 representative genes, individual knockdown of 27 genes caused PM-cholesterol content to decrease by >50% (Figure S1F). Fifteen genes exhibited markedly enhanced cholesterol accumulation in cells as shown by filipin staining (Figure S1G). These results confirmed the reliability of our screen.

Intriguingly, genes encoding peroxisomal proteins were statistically enriched (Figures 1D and 1E). When the peroxisomal hits, including *ABCD1*, *ACOT8*, *BAAT*, *TMEM135*, *PEX1*, *PEX3*, *PEX6*, *PEX10*, and *PEX26* were individually knocked down, the PM-cholesterol level significantly decreased by 35%–84% as compared to control (Figure 2A). Cholesterol accumulation was observed in lysosome, but not peroxisome (Figures 2B and S2B).

### Peroxisome Forms Transient and Dynamic Contacts with Lysosome

How can depletion of peroxisomal proteins lead to cholesterol accumulation in lysosome? To answer this question, we used *ABCD1*, a peroxisomal membrane protein and also one of the strongest hits from our screen, as a representative to investigate the mechanism.

*ABCD1* mainly colocalized with the peroxisome marker PMP70 as expected. However, significant amount of colocalization between *ABCD1* and lysosome marker LAMP1 was surprisingly observed (Figures 2C, 2D, and S3A). Using GM130 as marker for the Golgi apparatus and EEA1 and Rab5 as early endosome markers, we found the lysosome-peroxisome contact was very specific as there was little detectable association between peroxisome and these two organelles (Figures 2D and S3A). Is the apparent colocalization of lysosome and peroxisome due to the sporadic distribution of *ABCD1* in lysosome? SKL is a strong peroxisome localization signal and the EGFP-His<sub>6</sub>-SKL protein is widely used to label peroxisome. We analyzed other peroxisome markers such as transfected EGFP-His<sub>6</sub>-SKL and endogenous PMP70 to rule out potential interference of particular marker or antibody and found a similar partial colocalization between lysosome and peroxisome (Figures S3A–S3C).

We took extra caution to further validate this phenomenon using 3D reconstitution, super resolution structured illumination microscopy (SR-SIM), and electron microscopy. 3D reconstitution and high resolution confocal images showed that the small membrane interaction between lysosome and peroxisome could indeed be observed using different microscopic methods (Figures 2E and 2F). Moreover, lysosome and peroxisome formed contacts in primary mouse hepatocytes detected by transmission electron microscopy (Figure 2G). With these validations, we named this phenomenon lysosome-peroxisome membrane

contact (LPMC). To our knowledge, the LPMC has not been reported before.

Time-lapse microscopy was next employed to understand the LPMC dynamics in living cells. It revealed that the contact between lysosome and peroxisome was only transient. In a time frame of a few dozen to 100 s, a particular peroxisome formed a contact with one lysosome, then was released and moved away. It could then associated with another lysosome in a similar time frame (Figure 2H; Movies S1 and S2). Notably, we observed no fusion of lysosome with peroxisome (Figure 2H). Consistently, a lysosomal matrix protein such as NPC2 was not detected in peroxisome when LPMC formed (Figure S3B).

To further validate the LPMC, we designed an organelle co-precipitation assay (Figure S3D). The cells stably expressing EGFP-His<sub>6</sub>-SKL were lysed without disturbing organelle integrity, and the membrane fractions were incubated with Ni Sepharoses to pull down peroxisome. The isolated fractions were then examined by fluorescent images of Ni Sepharoses and western blot. As shown in Figure S3E, NPC1-mCherry-labeled lysosome (red) could be observed on the beads covered by peroxisome (green). On the other hand, mCherry-Rab5-labeled early endosome was not co-precipitated suggesting the LPMC was specific. In line with these results, western blot analysis showed that the lysosomal protein LAMP1 was efficiently co-precipitated with peroxisome, but markers for other organelles were not (Figure S3F). Together, these lines of evidences strongly demonstrate the presence of LPMC.

We next examined if LPMC is regulated. Knockdown of *NPC1* or *ABCD1* significantly decreased the LPMC, with this effect being evident using both cell imaging and organelle co-precipitation methods (Figures 3A–3C). Depletion of other peroxisomal functional proteins such as *PEX1* also led to less LPMC (Figures S2B and S2C). More importantly, the LPMC was significantly reduced under cholesterol depletion status, and this reduction could be time-dependently reversed by cholesterol repletion from LDL (Figures 3D–3F). Knockdown of *LDLR*, *Clathrin heavy chain* (*CHC*), or co-depleting of adaptor proteins genes including *AP2 subunit alpha 2*, *ARH*, and *Dab2* to inhibit LDL endocytosis not only attenuated lysosomal cholesterol replenishment, but also decreased the LPMC (Figures 3G and 3H). These results suggest that cellular cholesterol content regulates LPMC, which also requires proper functions of lysosome and peroxisome.

### Synaptotagmin VII Is a Lysosomal Protein Binding Peroxisome

We next sought to identify the molecules bridging LPMC. A multi-arm proteomics approach was employed to analyze lysosomal membrane proteins, peroxisomal proteins, and NPC1 interacting proteins (Figures S4A and S4B; Table S2). After merging of the protein lists, candidates involved in vesicle fusion or organelle dynamics were selected as refined candidates for

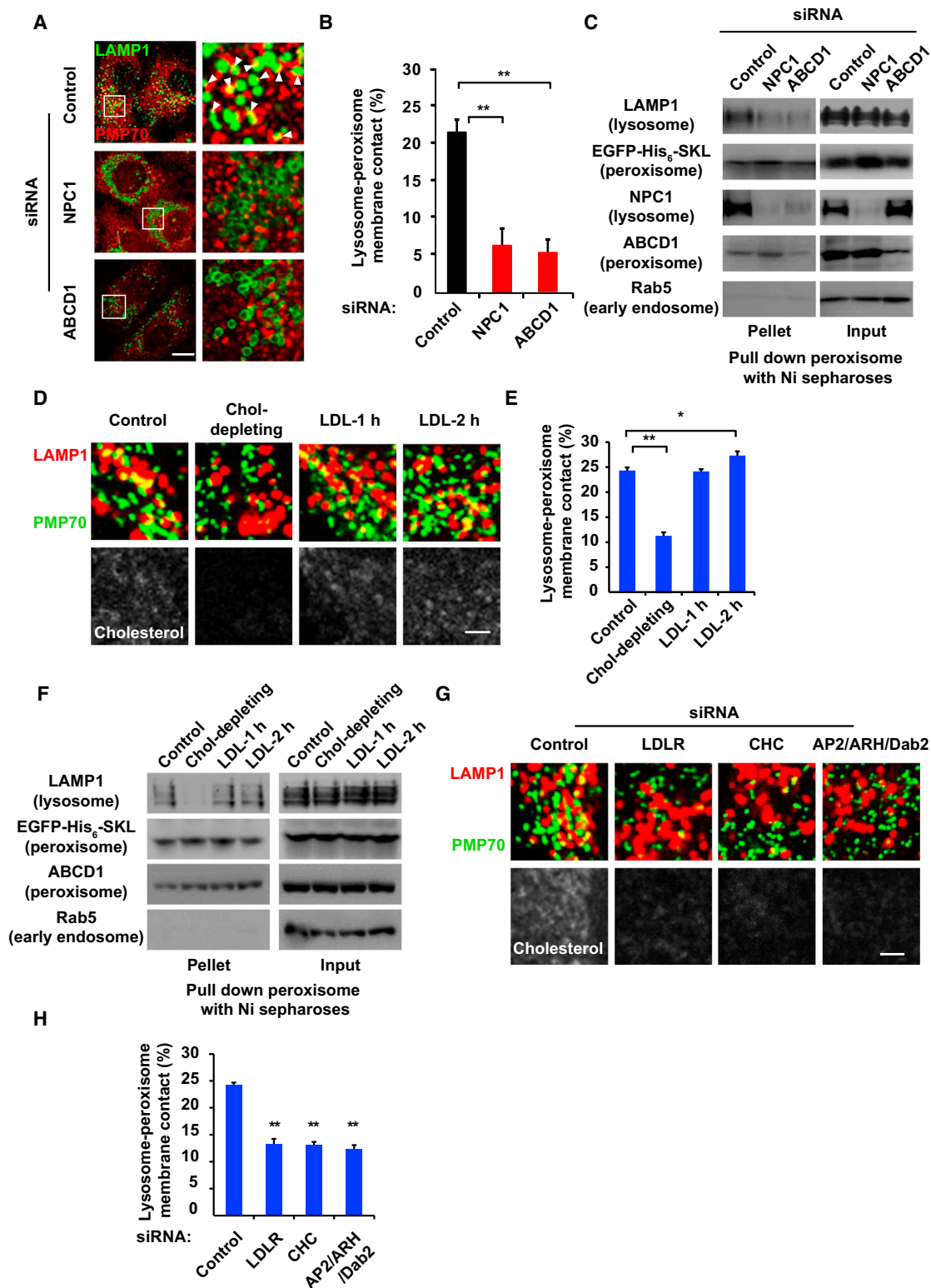
(E) A representative SR-SIM image of the overlaid endogenous LAMP1 (green) and PMP70 (red) images. Arrowheads indicate LPMC sites. Scale bar, 10  $\mu$ m.

(F) HeLa cells were immunostained with antibodies against LAMP1 and PMP70 and analyzed by Volocity-3D software. Arrowheads indicate LPMC sites. Scale bar, 10  $\mu$ m.

(G) Transmission electron micrograph of the LPMC in a mouse liver cell. L, lysosome, P, peroxisome. Scale bar, 500 nm.

(H) SV589 cells were transfected with EGFP-SKL and NPC1-mCherry. Time-lapse images were acquired. Scale bar, 500 nm. See also Movie S2.

See also Figures S2 and S3 and Movies S1 and S2.



**Figure 3. Regulation of Lysosome-Peroxisome Membrane Contacts**

(A) SV589 cells were transfected with indicated siRNAs and immunostained with antibodies against LAMP1 (green) and PMP70 (red). Scale bar, 10  $\mu$ m.

(B) Quantification of LPMC in (A). Data represent mean  $\pm$  SD. \*\* $p < 0.01$ , one-way ANOVA ( $n = 4$ , 35 cells per independent experiment).

(C) Lysosome-peroxisome association revealed by organelle co-precipitation assay.

(legend continued on next page)

RNAi validation (Table S2). Out of the 16 candidates, Synaptotagmin VII (Syt7) was the only that passed the validation: its knockdown but not that of others caused clear cholesterol accumulation in cells (Figure S4C).

Synaptotagmin is a family of proteins involving in vesicle interaction and fusion. Syt7 is widely expressed and plays important role in lysosomal exocytosis, membrane resealing, and wound healing (Andrews and Chakrabarti, 2005). Syt7 mainly colocalized with the lysosome marker LAMP1 (Figure 4A). Similarly to NPC1 and LAMP1, Syt7 significantly colocalized with the peroxisome marker PMP70 but not markers for Golgi or early endosome (Figures 4A and 4B). Knockdown of Syt7 resulted in cholesterol accumulation in lysosome (Figure 4C), and the LPMC was also dramatically diminished (Figure 4D). Syt7 is a transmembrane protein with a short N-terminal ectodomain, a single transmembrane segment, and a large cytosolic region containing two tandem  $\text{Ca}^{2+}$ -binding C2 domains (C2A and C2B, Figure 4E). The C2A and C2B domains are responsible for the  $\text{Ca}^{2+}$ -dependent interactions between Syt7 and SNAREs or phospholipids. When overexpressed, these domains compete for binding to SNAREs or phospholipids and function as dominant-negative forms (Desai et al., 2000). We utilized a similar method and found that overexpression of C2A or C2B domain dramatically inhibited LPMC in cell imaging and organelle coimmunoprecipitation (coIP) (Figures 4F, 4G, and S4D), accompanied by cholesterol accumulation in cells (Figure 4H).

We further developed an in vitro reconstitution assay to dissect the mechanism of LPMC (Figure S5A). Briefly, EGFP-His<sub>6</sub>-SKL-labeled peroxisome and NPC1-FLAG-mCherry-labeled lysosome were separately isolated by density gradient centrifugation. The peroxisomes were further precipitated by Ni Sepharoses and incubated with purified lysosome fractions. After incubation, Ni Sepharoses were separated by centrifugation and subjected to confocal microscopy and western blot. As shown in Figure S5B, lysosome labeled as red was pulled down with peroxisome in the presence of cytosol and ATP/GTP. Consistently, the lysosome marker NPC1-FLAG-mCherry was co-precipitated at this condition (Figure 4I). These results suggest energy and some cytosolic proteins may facilitate LPMC. Addition of dominant-negative Syt7-C2AB protein (Figure S5C) in the incubation step blocked the lysosome peroxisome interaction (Figures 4J and S5D). Similarly, when lysosomes from Syt7 or NPC1 RNAi-depleted cells were incubated with control peroxisome, the LPMC was significantly reduced. Conversely, when lysosome from control cells was incubated with peroxisome from Syt7 or NPC1 RNAi-depleted cells, the LPMC was not affected (Figures 4K and S5E). These findings indicate Syt7 is a lysosomal protein required for LPMC formation.

### PI(4,5)P<sub>2</sub> in Peroxisome Membrane Bridges LPMC

It has been documented that SNAREs mediate membrane contacts and fusion throughout the secretory pathway (Chen and Scheller, 2001; Weber et al., 1998). Organelles such as Golgi, ER, and lysosome are all maintained by SNARE-based fusion events. However, so far, no peroxisomal SNARE protein has been identified. Consistent with previous studies (Matsumoto et al., 2003), no SNARE family protein was identified in our peroxisomal proteomics (Table S2). Because Syt7 binds to phospholipids besides SNARE, we hypothesized that Syt7-mediated LPMC might be through its interaction with peroxisomal phospholipids. To test this hypothesis, we examined the binding specificity of Syt7-C2AB to various phospholipids in a PIP-strip assay. Syt7-C2AB mainly bound PI(4,5)P<sub>2</sub> and to a much lesser extent PI(5)P and PS; no signal was observed for other phospholipids (Figure 5A). It has been reported that peroxisome can synthesize significant amounts of PIP2 including PI(4,5)P<sub>2</sub> (Jeynov et al., 2006). To further validate Syt7-PI(4,5)P<sub>2</sub> interaction under a more relevant format, we performed the liposome flotation assay using liposomes mimicking phospholipid composition of the mammalian peroxisome membrane (PC:PE:PI:PS = 54:36:5:5) (Hardeman et al., 1990). As shown in Figure 5B, when mixed with blank liposomes or PI5P containing liposomes, the His<sub>6</sub>-C2AB protein was predominantly detected in the bottom fraction. Trace amount of His<sub>6</sub>-C2AB in middle and top fractions was also detected, possible due to the weak binding of C2AB to PS and PI5P. In contrast, the majority of His<sub>6</sub>-C2AB protein was co-floated with liposomes containing PI(4,5)P<sub>2</sub> to the top fraction. These results demonstrated that the C2AB domain of Syt7 interacts with PI(4,5)P<sub>2</sub> in membrane.

Next, we sought to determine whether the Syt7-PI(4,5)P<sub>2</sub> interaction functions to bridge LPMC using an inducible FKBP12-FRB heterodimerization system to deplete PI(4,5)P<sub>2</sub> on peroxisome (Figure 5C). In the constructed SV589 cells, FKBP12 was targeted to peroxisome by fusion with PEX-mCherry, and the inositol polyphosphate 5-phosphatase synaptotagmin 2 (SYNJ2) was kept in cytoplasm fused with mCitrine-FRB. Application of the chemical inducer rapamycin led to peroxisome membrane recruitment of mCitrine-FRB-SYNJ2 by binding PEX-mCherry-FKBP12 (Kapitein et al., 2010), which rapidly and irreversibly converted PI(4,5)P<sub>2</sub> to PI(4)P (Figures 5C and 5D). As shown in Figures 5E and 5F, rapamycin treatment caused a significant decrease of LPMC and cellular cholesterol accumulation. The cell expressing only mCitrine-FRB was a control showing no change of LPMC or cholesterol aggregation. Although cellular PI(4,5)P<sub>2</sub> also presents on PM, depletion of PI(4,5)P<sub>2</sub> in PM by a similar strategy did not decrease LPMC or cause cholesterol accumulation (Figures S6A–S6C). Furthermore, anti-PI(4,5)P<sub>2</sub> antibody specifically reduced the association between lysosome and peroxisome in vitro (Figures 5G, 5H, and S5F). Together,

(D) HeLa cells were incubated in cholesterol-depleting medium for 16 hr and then refed with LDL for different time durations. Cells were stained with filipin (gray) and antibodies against LAMP1 (red) and PMP70 (green). Scale bar, 2  $\mu\text{m}$ .

(E) Quantification of LPMC in (D). Data represent mean  $\pm$  SD ( $n = 4$ , 35 cells per independent experiment). \*\* $p < 0.01$ , \* $p < 0.05$ .

(F) Organelle co-precipitation assay was performed to validate LPMC when cells were grown under conditions shown in (D).

(G) SV589 cells transfected with indicated siRNAs were stained with filipin (gray) and antibodies against LAMP1 (red) and PMP70 (green). Scale bar, 2  $\mu\text{m}$ .

(H) Quantification of LPMC in (G). Data represent mean  $\pm$  SD ( $n = 4$ , 35 cells per independent experiment). \*\* $p < 0.01$ .

See also Figure S3.



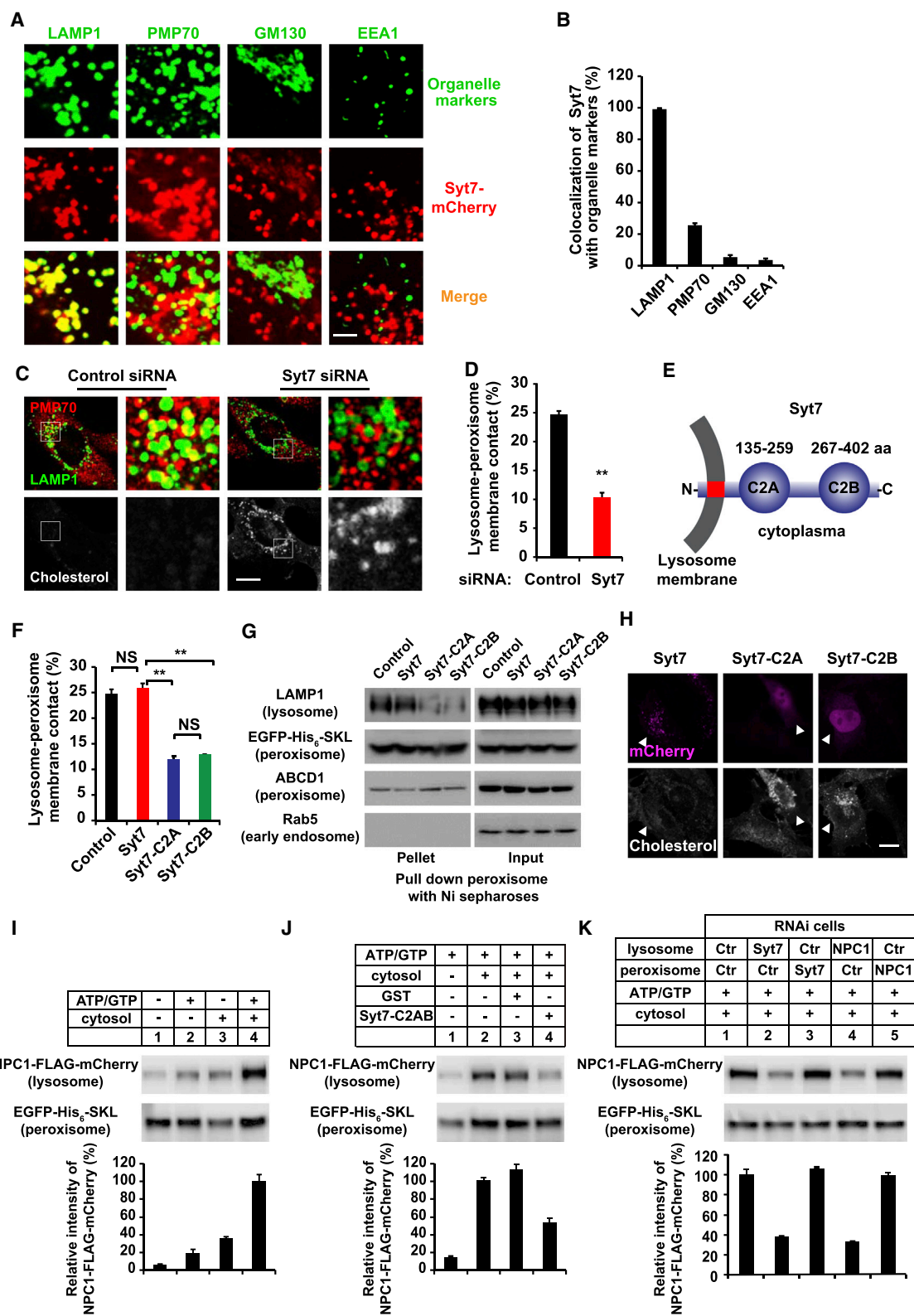


Figure 4. Synaptotagmin VII Is a Lysosomal Protein Bridging LPMC

(A) SV589 cells transfected with Syt7-mCherry were assessed by immunostaining with indicated antibodies. Scale bar, 2  $\mu$ m.

(B) Quantification of Syt7 colocalization with organelle-specific markers. Data represent mean  $\pm$  SD (n = 4, 35 cells per independent experiment).

(legend continued on next page)

these data demonstrate that PI(4,5)P<sub>2</sub> in peroxisome membrane is required for LPMC and proper cholesterol transport.

Because PI(4,5)P<sub>2</sub> is critical for LPMC, we reasoned that the peroxisome genes from our screen might affect peroxisomal PI(4,5)P<sub>2</sub> level either directly or indirectly. Indeed, a pronounced decrease in the amount of PI(4,5)P<sub>2</sub> in peroxisomal lipid extraction was detected using dot blots with anti-PI(4,5)P<sub>2</sub> antibody after knocking down *ABCD1* or other peroxisomal hits (Figure S6D). These data suggest that the nine peroxisome proteins may not directly bind Syt7 but rather influence peroxisomal PI(4,5)P<sub>2</sub> level thereby affecting lysosome association.

### Cholesterol Transport through LPMC

To monitor cholesterol transport directly, we used <sup>3</sup>H-cholesterol in the in vitro reconstitution assay (Figure 6A). Briefly, <sup>3</sup>H-cholesterol-labeled lysosome was isolated by density centrifugation from HEK293T cell pre-incubated with <sup>3</sup>H-cholesterol. Peroxisome was purified from unlabeled cells. The lysosome and peroxisome were then applied to the in vitro reconstitution system. After incubation, EGTA washing was performed to dissociate lysosome from peroxisome while leaving the peroxisome on Ni Sepharoses. The <sup>3</sup>H-cholesterol in peroxisome was then measured. To control the specificity, antibodies against PI(4,5)P<sub>2</sub> or unrelated IgG were applied. The <sup>3</sup>H-cholesterol in peroxisome increased in a time-dependent manner and this increase was blocked by anti-PI(4,5)P<sub>2</sub> antibody (Figures 6B and S7A). In addition, lysosomes prepared from *NPC1* or *Syt7* RNAi cells failed to support cholesterol transfer to peroxisome (Figure 6C), because LPMC did not form when *NPC1* or *Syt7* was depleted from lysosome (Figures 4K and S5E). These data demonstrate that cholesterol can transfer from lysosome to peroxisome depending on LPMC in vitro.

What about in cells? We performed confocal microscopy on HeLa cells refed with LDL and observed a time-dependent increase of co-localization between peroxisome and cholesterol (Figures S7B and S7C). We also directly measured the cholesterol level in isolated lysosome and peroxisome after incubation with <sup>3</sup>H-cholesteryl oleate containing LDL (scheme in Figure S7D). The lysosome and peroxisome were both labeled by <sup>3</sup>H-cholesterol although peroxisome label was less (Figure 6D). Results from western blot of organelle markers excluded the contamination with other organelles (Figure S7E). Furthermore,

knockdown of *NPC1* or *ABCD1* caused significant increase of <sup>3</sup>H-cholesterol in lysosome and decrease in peroxisome (Figure 6D). LDL pulse chase experiment (Figure S7F) followed by SR-SIM microscopy showed that there was overlay of peroxisome with cholesterol-loaded lysosome, or cholesterol (Figure S7G). These data suggest cholesterol flows from lysosome to peroxisome in cells.

To further investigate whether LPMC is required for LDL-cholesterol transport to the ER, we performed SREBP cleavage and cholesterol esterification assays because it is well established that cholesterol derived from LDL prevents SREBP processing and stimulates cholesterol esterification once it reaches the ER. The results showed that LDL-cholesterol could efficiently block SREBP processing (Figure 6E) and stimulate cholesterol esterification (Figure 6F) in control cells. However, these effects were markedly blunted in *NPC1*, *ABCD1*, or *Syt7* RNAi cells (Figures 6E and 6F); demonstrating that cholesterol transport to ER was largely impaired when LPMC was disrupted.

With the current data and information from previous reports (Kwon et al., 2009), we propose the below model for cholesterol transport from lysosome to peroxisome. After internalization, LDL particles are delivered to lysosome where LDL-containing cholesteryl ester is hydrolyzed to unesterified cholesterol. The luminal NPC2 protein binds free cholesterol with the 8-carbon isooctyl side chain buried within the binding pocket and hands over the cholesterol molecule to the N-terminal domain of NPC1. The NPC1-N-terminal domain can penetrate the glycocalyx and facilitate cholesterol to insert into the lysosomal membrane. Lysosome and peroxisome form close membrane contacts through interaction between Syt7 and PI(4,5)P<sub>2</sub>. Thus, cholesterol can move from lysosome to peroxisome (Figure 6G).

### Intracellular Cholesterol Accumulation in Peroxisomal Disorders

*ABCD1* mutation causes X-ALD, which is a neurological disease with progressive CNS demyelination and adrenal insufficiency (Forss-Petter et al., 1997). X-ALD is one of the prevalent peroxisomal disorders and there is no effective treatment (Moser et al., 2005). Our work has demonstrated cholesterol transports from lysosome to peroxisome through LPMC, and *ABCD1* depletion impairs LPMC and leads to cholesterol accumulation. However, there is no previous report on cholesterol transport defect in

(C) SV589 cells transfected with indicated siRNAs were stained with filipin (gray) and antibodies against LAMP1 (green) and PMP70 (red). Insets show high magnification of the areas framed by a white box. Scale bar, 10  $\mu$ m.

(D) Quantification of LPMC in (C). Data represent mean  $\pm$  SD (n = 4, 35 cells per independent experiment). \*\*p < 0.01.

(E) Domain structure of the Syt7 protein.

(F) SV589 cells transfected with mCherry, Syt7, C2A, or C2B of Syt7 were assessed by immunostaining with antibodies against LAMP1 and PMP70. Shown is the quantification of LPMC. Data represent mean  $\pm$  SD (n = 4, 30 cells per independent experiment). NS, not significant, \*\*p < 0.01. The fluorescence images of cells are shown in Figure S4D.

(G) HeLa/EGFP-His<sub>6</sub>-SKL cells were transfected with the indicated plasmids and the lysosome-peroxisome association was analyzed by organelle co-precipitation assay.

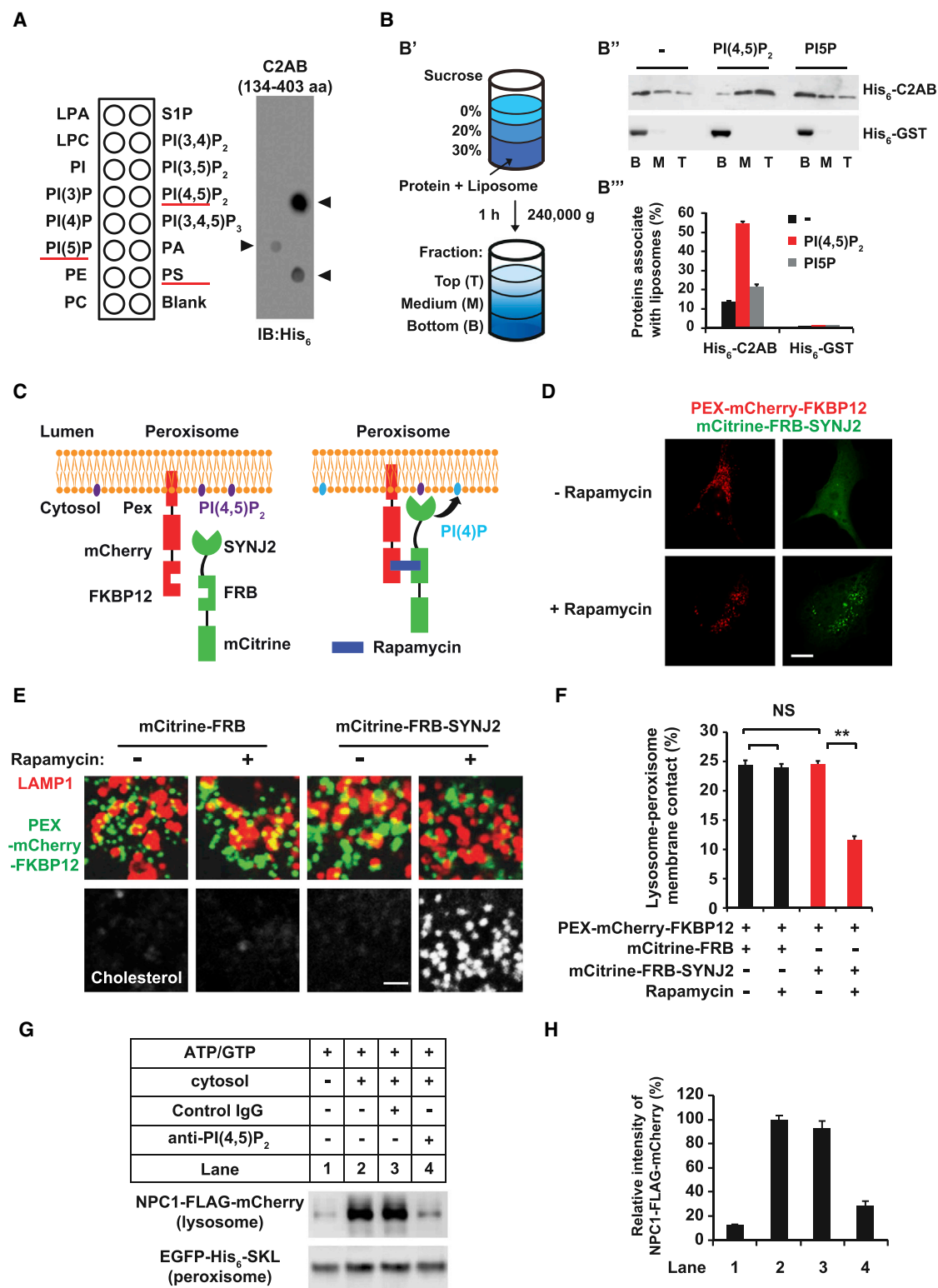
(H) SV589 cells transfected with the indicated plasmids were stained with filipin (gray). Arrowheads indicate the cells expressing Syt7 or Syt7 variants (magenta). Scale bar, 10  $\mu$ m.

(I) In vitro reconstitution of LPMC. The images of Ni Sepharoses are shown in Figure S5B.

(J) Recombinant GST or Syt7-C2AB protein was applied in the in vitro reconstitution system. The images of Ni Sepharoses are shown in Figure S5D.

(K) Lysosome or peroxisome was purified from cells transfected with indicated siRNAs and then used for the in vitro reconstitution assay. The images of Ni Sepharoses are shown in Figure S5E. Ctr, control.

See also Figures S4 and S5 and Table S2.



**Figure 5. PI(4,5)P<sub>2</sub> of Peroxisome Is Required for LPMC**  
(A) Protein-lipid overlay. A scheme of the PIP-strip membrane is shown (left). Arrowheads indicate specific lipids binding. Red lines highlight the phospholipid species.

(legend continued on next page)

X-ALD or any other peroxisomal disorders. Therefore, we sought to validate our findings in vivo by examining if there is cholesterol accumulation in *ABCD1* knockout (KO) animal models and fibroblasts of human patients with different types of peroxisomal disorders.

As shown in Figure 7A, cholesterol accumulated in zebrafish embryo cells injected with morpholino antisense oligomer (MO) against *NPC1* or *ABCD1*. Furthermore, cholesterol accumulation was observed in fibroblasts, cerebellum, and adrenal gland of *ABCD1* KO mice (Figures 7B and 7C), a well-accepted animal model capturing the pathological characteristics of X-ALD. Interestingly, in the adrenal gland cholesterol deposits were located almost exclusively in the cortex but not in the medulla (Figure 7C), correlating with *ABCD1*'s specific expression in the cortex (Troffer-Charlier et al., 1998).

Because it is known that the *ABCD1* KO mice do not show an abnormal behavioral or neurological phenotype up to 15 months, we analyzed the behavioral deficits associated with CNS demyelination using rotarod test at the age of 7 and 20 months, respectively. When compared with WT littermates, the 20-month-old *ABCD1* KO mice displayed a marked impairment (19%) in their ability to stay on top of a rotated cylinder during 2 days trial, while the 7-month-old *ABCD1* KO mice were not affected (Figure 7D). Open field mobility paradigm was also used to study spontaneous locomotion and exploratory behavior. As shown in Figures 7E and 7F, the 20-month-old *ABCD1* KO mice exhibited significantly fewer numbers of rearings and traveled shorter distances in comparison with WT mice or 7-month-old *ABCD1* KO mice. This is important because the cholesterol accumulation occurs as early as 7 months (Figure 7C), long before the manifestation of the neurological phenotypes (20-month-old), suggesting not only that losing of *ABCD1* leads to cholesterol trafficking defects, but also that intracellular cholesterol accumulation might be a mechanism causing X-ALD symptoms.

To further evaluate the role of peroxisome in cholesterol trafficking, cultured fibroblasts from patients with X-ALD, or two peroxisome biogenesis disorders Infantile Refsum disease (IRD) and Zellweger syndrome (ZS) were used for cholesterol staining. As shown in Figure 7G, drastic cholesterol accumulation was observed in these fibroblasts, suggesting peroxisome plays an essential role in intracellular cholesterol transport.

## DISCUSSION

Using an elegantly designed cellular system, our genome-wide shRNA screen allows a comprehensive dissection of the genes

and pathways that may regulate intracellular cholesterol transport. Besides the previously known cholesterol transport gene like *NPC1*, we uncovered over 300 additional genes, among which the genes encoding peroxisomal proteins were highly enriched.

We showed that peroxisome played an essential role in intracellular cholesterol transport through forming membrane contacts with lysosome. We provided multiple lines of evidence to solidify this observation. First, LPMC was observed by confocal microscopy and decreased by cholesterol depletion and knocking down of *NPC1* or *ABCD1*. Second, super resolution microscopy showed the overlapping signals between peroxisome and lysosome (Figure 2E). Third, 3D reconstitution verified LPMC from different angles (Figure 2F). Fourth, transmission electron micrographs directly observed LPMC in primary mouse hepatocytes (Figure 2G). Fifth, time-lapse imaging showed the LPMC is dynamic in living cells (Figure 2H). Sixth, organelle co-precipitation assay detected the physical interaction between peroxisomes and lysosomes (Figures 3C and S3F). Seventh, in vitro reconstitution assay confirmed that lysosome and peroxisome can form contacts specifically (Figures 4I and S5).

As for the molecules bridging LPMC, our data demonstrate lysosomal protein Syt7 binds peroxisomal lipid PI(4,5)P<sub>2</sub> to form a transient contact. How are Syt7 activation and peroxisomal PI(4,5)P<sub>2</sub> level regulated? It is well known that calcium can bind Syt7 leading to a conformational change (Fukuda and Mikoshiba, 2001). Meanwhile, the level of PI(4,5)P<sub>2</sub> can be modulated by phosphatidylinositol kinases and phosphatases. Its distribution is also under dynamic regulation. Therefore, how different proteins/pathways regulate Syt7 and PI(4,5)P<sub>2</sub> and then influence LPMC and cholesterol transport is a particular interesting subject for further exploration. The in vitro reconstitution assay developed in this study would be a powerful tool. Our screen discovered 9 peroxisomal proteins including *ABCD1*, knockdown of which individually leads to lowered peroxisomal PI(4,5)P<sub>2</sub> level (Figure S6D). These nine peroxisomal proteins cover different functions and are all required for proper peroxisomal function. Therefore, the dysfunction of peroxisome may underlie the decrease of PI(4,5)P<sub>2</sub> and LPMC. Further studies are still needed to understand how these peroxisome proteins are functionally connected to PI(4,5)P<sub>2</sub> regulation.

Previous studies have indicated that cholesterol can leave lysosome by vesicular or non-vesicular transport. Urano et al. (2008) showed that LDL-cholesterol can be transported from L/L to the trans-Golgi network through vesicular trafficking. Du et al. (2011) reported that ORP5, an oxysterol-binding protein-related

(B) B': workflow of the liposome flotation assay. B'': the presence of recombinant proteins in the top (T), middle (M), and bottom (B) fractions were detected by western blot using anti-His<sub>6</sub> antibody. B''': semiquantitative densitometric analysis of western blot in B''. The amount of liposomes-associated proteins was determined by comparing proteins present in the top fraction to the total amount of proteins present in the top, middle, and bottom fractions.

(C) Schematic representation of the rapamycin-inducible heterodimerization system used to recruit SYNJ2 to the peroxisome membrane.

(D) Validation of the rapamycin-inducible system in SV589 cells. Scale bar, 10  $\mu$ m.

(E) SV589 cells were transfected with PEX-mCherry-FKBP12 together with either mCitrine-FRB or mCitrine-FRB-SYNJ2. Cells were then treated with rapamycin, stained with filipin (gray), and immunostained with antibody against LAMP1, followed by Cy5-conjugated anti-mouse secondary antibody (pseudocolor, red). Scale bar, 2  $\mu$ m.

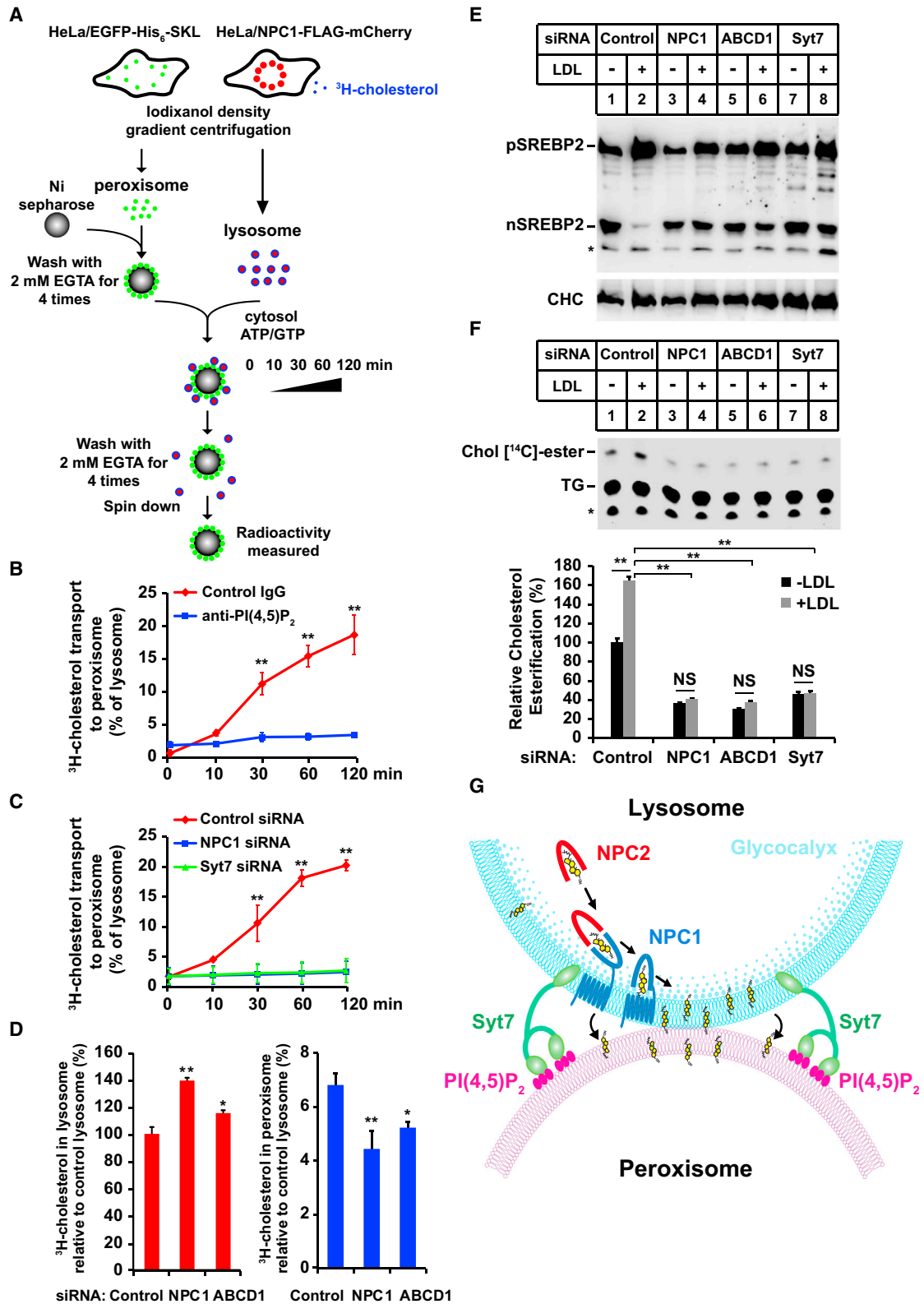
(F) Quantification of LPMC in (E). Data represent mean  $\pm$  SD (n = 4, 30 cells per independent experiment). NS, not significant, \*\*p < 0.01.

(G) Anti-PI(4,5)P<sub>2</sub> or control IgG was applied in the in vitro reconstitution system. The images of Ni Sepharoses are shown in Figure S5F.

(H) Semiquantitative densitometric analyses of (G).

See also Figures S5 and S6.





(legend on next page)

protein, may mediate cholesterol efflux from lysosome to ER through binding cholesterol and NPC1. Here, cholesterol transport across LPMC is another mechanism for cholesterol efflux from lysosome. Disruption of LPMC by different means causes significant lysosomal cholesterol accumulation. X-ALD animal models and fibroblasts of human patients with different types of peroxisomal disorders displayed drastic cholesterol accumulation (Figure 7), suggesting LPMC is a major route for cholesterol to leave the lysosomal membrane. Our in vitro reconstitution assay suggests cytosol may facilitate cholesterol movement from lysosome to peroxisome (Figure 4I). Finally, it is possible that cytosolic cholesterol binding proteins such as StarD4 and ORPs may accelerate the cholesterol movement from lysosomal membrane to peroxisome when LPMC forms.

After reaching peroxisome, the cholesterol might be further oxidized or participate in bile acid synthesis in peroxisome. Cholesterol is also required for peroxisome lipid raft assembly and peroxisome biogenesis (van der Zand and Tabak, 2013; Woudenberg et al., 2010), and we estimated peroxisome contains ~5% of total cellular cholesterol (data not shown). Because disrupting LPMC decreases PM cholesterol level (Figure 2A) and impairs LDL-cholesterol reaching the ER (Figures 6E and 6F), it is likely that peroxisome may associate with other organelles and deliver cholesterol to them. This notion is further supported by the observation that cholesterol in lysosome increased by 20%–40% whereas cholesterol in peroxisome only decreased by ~2% after LPMC disruption in cells (Figure 6D). Alternatively, cholesterol transport via LPMC may be tightly coupled with cholesterol modification including oxidation and esterification. It is interesting to further study how cholesterol transportation is affected by cholesterol modification and vice versa. Besides cholesterol transfer, LPMC may regulate other functions of lysosome and peroxisome, such as autophagy, mTOR signaling, and peroxisome biogenesis.

Dramatic cholesterol accumulation was observed in X-ALD animal models and human patients' fibroblasts with mutations in different peroxisomal genes (Figure 7). Notably, the cholesterol accumulation (7-month-old) occurs long before the manifestation of the neurological phenotypes (20-month-old), suggesting intracellular cholesterol accumulation might be a potential mechanism causing X-ALD symptoms. It was also noted that although there was early onset of very long chain fatty

acid accumulation, relief of its accumulation did not significantly improve the disease symptoms (Prieto Tenreiro et al., 2013). On the other hand, it was well established that the accumulation of cholesterol in NPC disease patients is the cause of neuron death and neurological phenotypes. Mobilizing cholesterol by cyclo-dextrin constitutes a beneficial treatment for NPC patients (Liu et al., 2010). Therefore, the cholesterol trafficking blockage may underlie the pathological mechanism of peroxisome disorders, which could provide novel strategies for diagnosis and treatment of these diseases.

In summary, through functional genome-wide RNAi screen and hits analysis, we demonstrate the existence of lysosome-peroxisome membrane contacts mediated by Syt7- PI(4,5)P<sub>2</sub> binding, through which cholesterol is transported from lysosome to peroxisome. Peroxisomal disorders display significant intracellular cholesterol accumulation prior to neuronal symptoms. Together, this study suggests a central role of peroxisome in intracellular cholesterol trafficking and highlights the clinical relevance of cholesterol transport in peroxisomal disorders.

## EXPERIMENTAL PROCEDURES

Materials and plasmids, cell culture, growth assay, liposome flotation assay, and other procedures are described in the [Extended Experimental Procedures](#).

### shRNA Screen and Analysis

HeLa cells were infected with the MISSION LentiPlex human pooled shRNA library consists of over 75,000 shRNA constructs from the TRC collection targeting 15,000+ human genes. Infected cells were selected with puromycin (2 µg/ml) for 4 days. After five rounds of AmB selections ([Extended Experimental Procedures](#)), survived populations were collected, and shRNA inserts were amplified from genomic DNA by PCR. PCR products were sequenced by deep-sequencing. All the deep sequencing data were log10 transformed and normalized to standard derivation from the screen-wide mean, which depicted as Z score [ $Z = (\text{gene's deep sequencing score} - \text{average deep sequencing score}) / \text{screen standard derivation}$ ]. Z score equal to 1.96 ( $p = 0.05$ ) was used as cut-off value to determine the screen hits. Genes with Z score over 1.96 ( $p < 0.05$ ) or targeted by five independent shRNAs were considered as screen hits.

### Organelle Co-Precipitation Assay

Triplicate samples for each treatment were homogenized in extraction buffer (5 mM MOPS [pH 7.65], with 0.25 M sucrose, 1 mM EDTA, 0.1% ethanol and protease inhibitors) and centrifuged at 1,000 × g for 10 min. Supernatants

## Figure 6. Transfer of <sup>3</sup>H-Cholesterol from Lysosome to Peroxisome

(A) Outline of the in vitro <sup>3</sup>H-cholesterol transfer assay.

(B) Ni Sepharoses bound-peroxisome was preincubated with anti-PI(4,5)P<sub>2</sub> or control IgG and then used for the in vitro <sup>3</sup>H-cholesterol transfer assay. Values are expressed as the percentage of <sup>3</sup>H-cholesterol in lysosome prior to reaction and presented as the mean ± SD of three independent repeats of experiments. \*\* $p < 0.01$ .

(C) Radiolabeled lysosomes were isolated from cells transfected with indicated siRNAs. Peroxisome was purified from wild-type cells and were then used for the in vitro <sup>3</sup>H-cholesterol transfer assay as in (A). Data are presented as the mean ± SD of three independent repeats of the experiments. \*\* $p < 0.01$ .

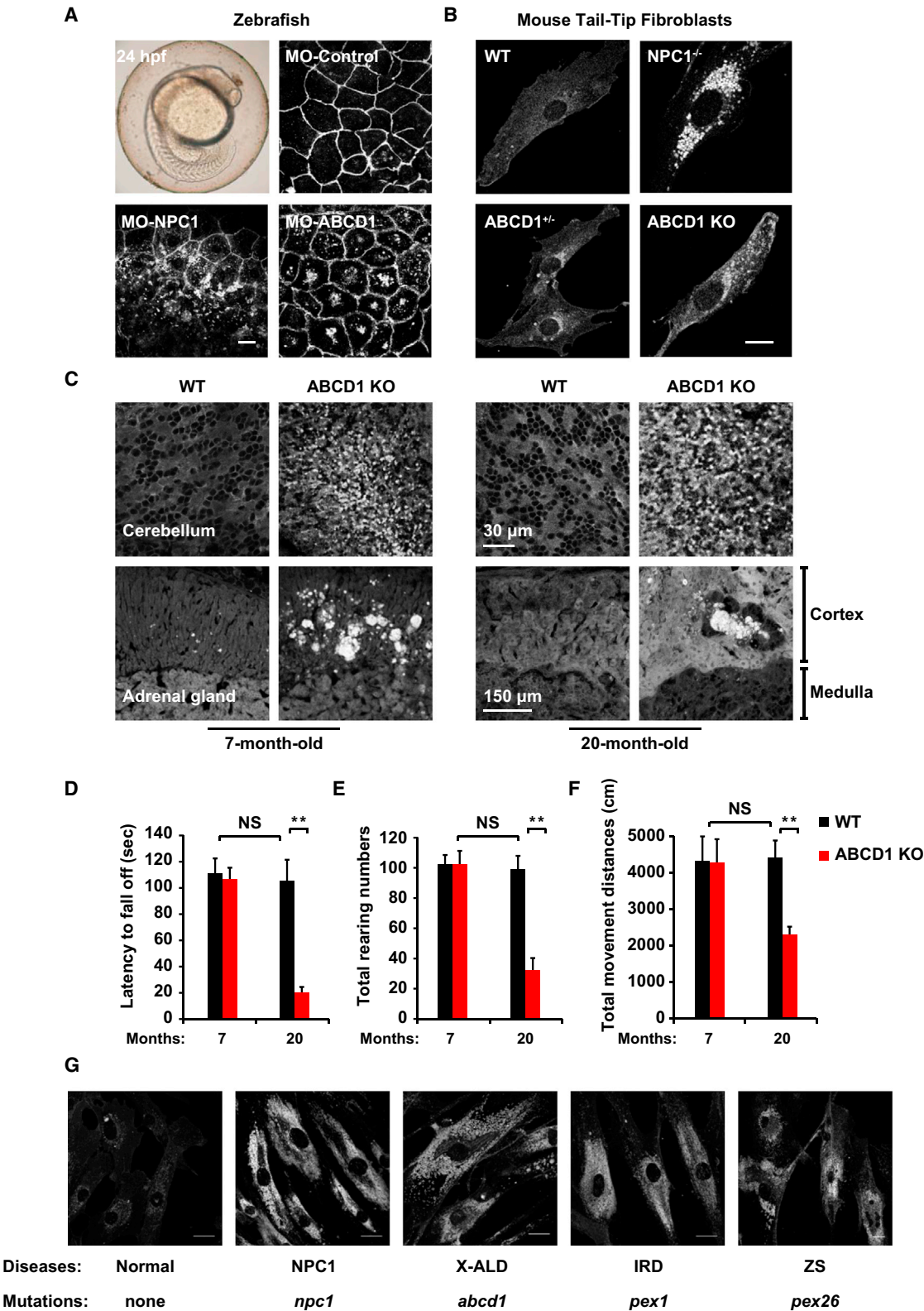
(D) HEK293T cells transfected with indicated siRNAs were depleted of cholesterol and then pulsed with <sup>3</sup>H-cholesteryl oleate-LDL for 3 hr. Then, lysosome and peroxisome were purified separately and the <sup>3</sup>H-cholesterol were measured. Values are expressed as percentage of control lysosome and presented as the mean ± SD of three independent experiments. \* $p < 0.05$ , \*\* $p < 0.01$ .

(E) HeLa cells transfected with indicated siRNAs were subjected to analysis of SREBP-2 cleavage. pSREBP2, precursor of SREBP2; nSREBP2, nuclear form of SREBP2; CHC, clathrin heavy chain. \*Indicates the nonspecific band.

(F) HeLa cells transfected with the indicated siRNAs were subjected to cholesterol esterification assay. TG, triacylglycerol. Quantification of cholesteryl [<sup>14</sup>C]-esters was analyzed by Image J. NS, not significant, \*\* $p < 0.01$ .

(G) A working mechanism of LDL-derived cholesterol transport out of lysosome.

See also [Figure S7](#).



**Figure 7. Cholesterol Accumulation in Animals and Human Patients with Peroxisomal Disorders**

(A) Filipin staining of unesterified cholesterol in zebrafish embryos. Scale bar, 10  $\mu$ m.  
(B) Filipin staining of the tail-tip fibroblast cells from the mice at the age of 7 months (n = 4 per group). Scale bar, 10  $\mu$ m.

(legend continued on next page)

were incubated with Ni Sepharoses at 4°C rotating for 2 hr. Beads were washed five times with extraction buffer. Then, 1  $\mu$ l Ni Sepharoses were mounted and analyzed by confocal microscope. Proteins bound to Sepharoses were eluted and subjected to western blot.

#### In Vitro Reconstitution Assay

EGFP-His<sub>6</sub>-SKL-labeled peroxisome and NPC1-FLAG-mCherry-labeled lysosome were first isolated by iodixanol density gradient centrifugation, respectively. The lysosome fractions were diluted with reconstitution buffer (250 mM sucrose, 1 mM DTT, 1 mM MgCl<sub>2</sub>, 50 mM KCl, and 20 mM HEPES [pH 7.2]), precipitated at 28,000  $\times$  g for 30 min and resuspended in reconstitution buffer. The peroxisome fractions were incubated with Ni Sepharoses, washed with reconstitution buffer plus 2 mM EGTA for four times, and then incubated with lysosome in the presence or absence of 1 mg/ml cytosol, 1 mM ATP, 1 mM GTP, and ATP-regenerating system (30 mM creatine phosphate, 0.05 mg/ml creatine kinase) at 37°C for 30 min. If needed, anti-PI(4,5)P<sub>2</sub> and Syt7-C2AB protein were applied 10 min at 4°C before the addition of ATP-regenerating system and cytosol. Ni Sepharoses were spun down, washed with reconstitution buffer, and subjected to microscopy and western blot.

#### In Vitro <sup>3</sup>H-Cholesterol Transfer Assay

Cells were incubated with 1  $\mu$ Ci/ml <sup>3</sup>H-cholesterol in growth medium overnight. The cells were then washed with PBS containing 0.2% BSA twice. The <sup>3</sup>H-cholesterol-labeled lysosome was isolated and its content of <sup>3</sup>H-cholesterol was measured using liquid scintillation. Peroxisome was purified from unlabeled cells by aforementioned method. The peroxisome and <sup>3</sup>H-cholesterol-labeled lysosome were subjected to in vitro reconstitution assay as described above. After incubation for different time durations at 37°C, the samples were spun down and lysosome was washed off by washing with reconstitution buffer plus 2 mM EGTA for four times. The <sup>3</sup>H-cholesterol on Ni Sepharoses bound-peroxisome was measured by liquid scintillation. The <sup>3</sup>H-cholesterol content in peroxisome was then normalized to the total input <sup>3</sup>H-cholesterol content of lysosome prior to reaction and was expressed as a percentage.

#### Measurement of LDL-Derived <sup>3</sup>H-Cholesterol in Lysosome and Peroxisome

After transfected with the indicated siRNAs, HEK293T cells were cultured in cholesterol-depleting medium for 16 hr and incubated with <sup>3</sup>H-cholesteryl oleate-LDL for 3 hr at 37°C. Then the cells were washed with PBS containing 0.2% BSA twice. Lysosome or peroxisome fractions were isolated by density gradient centrifugation separately and analyzed in a liquid scintillation counting.

#### Animals and Treatment

All animals were maintained and used in accordance with the guidelines of the Institutional Animal Care and Use Committee of the Shanghai Institutes for Biological Sciences. Mice were treated as described in the figure legends.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, two tables, and two movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.019>.

#### AUTHOR CONTRIBUTIONS

B.-L.S. conceived the project. B.-B.C., Y.-C.L., W.Q., B.-L.L., and B.-L.S. designed the experiments. B.-B.C. and Y.-C.L. performed the main experiments and analyzed data. C.X. performed SREBP cleavage and cholesterol esterifi-

cation assays. X.D. and H.Y. contributed to human patients' cells study. J.W. performed murine NPC1-TAP purification and mass spectrometry. Experiments were assisted by contributions from H.-H.M. Manuscript was written by B.-B.C., Y.-C.L., W.Q., and B.-L.S. with input from all the other authors.

#### ACKNOWLEDGMENTS

We thank Yu-Xiu Qu, Jie Xu, and Jie Qin for technical assistance. We thank Dr. Dawei Zhang (University of Alberta) for measuring ABCD1 activity, Dr. Fei Sun (Institute of Biophysics, Chinese Academy of Sciences) for helping with EM, Dr. TY Chang (Dartmouth Medical School) for CT43 cells, and Dr. Xiangdong Fu (Wuhan University) for helpful discussion and critical reading the manuscript. This work was supported by the grants from the National Natural Science Foundation (NNSF) of China (31430044, 31230020, and 91413112), the Ministry of Science and Technology (MOST) of China (2011CB910900 and 2012CB524900), Shanghai Science and Technology Committee (13XD1404100) and the 10,000 Talents Plan. X.D. and H.Y. were supported by grant APP1041301 from the National Health and Medical Research Council (NHMRC), Australia. H.Y. is a Senior Research Fellow of the NHMRC.

Received: August 11, 2014

Revised: December 18, 2014

Accepted: February 3, 2015

Published: April 9, 2015

#### REFERENCES

- Andreoli, T.E. (1973). On the anatomy of amphotericin B-cholesterol pores in lipid bilayer membranes. *Kidney Int.* 4, 337–345.
- Andrews, N.W., and Chakrabarti, S. (2005). There's more to life than neurotransmission: the regulation of exocytosis by synaptotagmin VII. *Trends Cell Biol.* 15, 626–631.
- Brown, M.S., and Goldstein, J.L. (1986). A receptor-mediated pathway for cholesterol homeostasis. *Science* 232, 34–47.
- Brown, M.S., and Goldstein, J.L. (1997). The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* 89, 331–340.
- Carstea, E.D., Morris, J.A., Coleman, K.G., Loftus, S.K., Zhang, D., Cummings, C., Gu, J., Rosenfeld, M.A., Pavan, W.J., Krizman, D.B., et al. (1997). Niemann-Pick C1 disease gene: homology to mediators of cholesterol homeostasis. *Science* 277, 228–231.
- Chang, T.Y., Chang, C.C., and Cheng, D. (1997). Acyl-coenzyme A:cholesterol acyltransferase. *Annu. Rev. Biochem.* 66, 613–638.
- Chang, T.Y., Chang, C.C., Ohgami, N., and Yamauchi, Y. (2006). Cholesterol sensing, trafficking, and esterification. *Annu. Rev. Cell Dev. Biol.* 22, 129–157.
- Chen, Y.A., and Scheller, R.H. (2001). SNARE-mediated membrane fusion. *Nat. Rev. Mol. Cell Biol.* 2, 98–106.
- Conner, S.D., and Schmid, S.L. (2002). Identification of an adaptor-associated kinase, AAK1, as a regulator of clathrin-mediated endocytosis. *J. Cell Biol.* 156, 921–929.
- Desai, R.C., Vyas, B., Earles, C.A., Littleton, J.T., Kowalchuck, J.A., Martin, T.F., and Chapman, E.R. (2000). The C2B domain of synaptotagmin is a Ca(2+)-sensing module essential for exocytosis. *J. Cell Biol.* 150, 1125–1136.
- Du, X., Kumar, J., Ferguson, C., Schulz, T.A., Ong, Y.S., Hong, W., Prinz, W.A., Parton, R.G., Brown, A.J., and Yang, H. (2011). A role for oxysterol-binding protein-related protein 5 in endosomal cholesterol trafficking. *J. Cell Biol.* 192, 121–135.

(C) Filipin staining of cerebellum and adrenal gland from different genotypes of mice.

(D) Motor performance of mice in a rotarod test.

(E) Quantification of rearing behavior of mice submitted to the open field test (n = 4 per group).

(F) The total distance moved within the open field arena (40 cm  $\times$  40 cm) was assessed over 15 min (n = 4 per group).

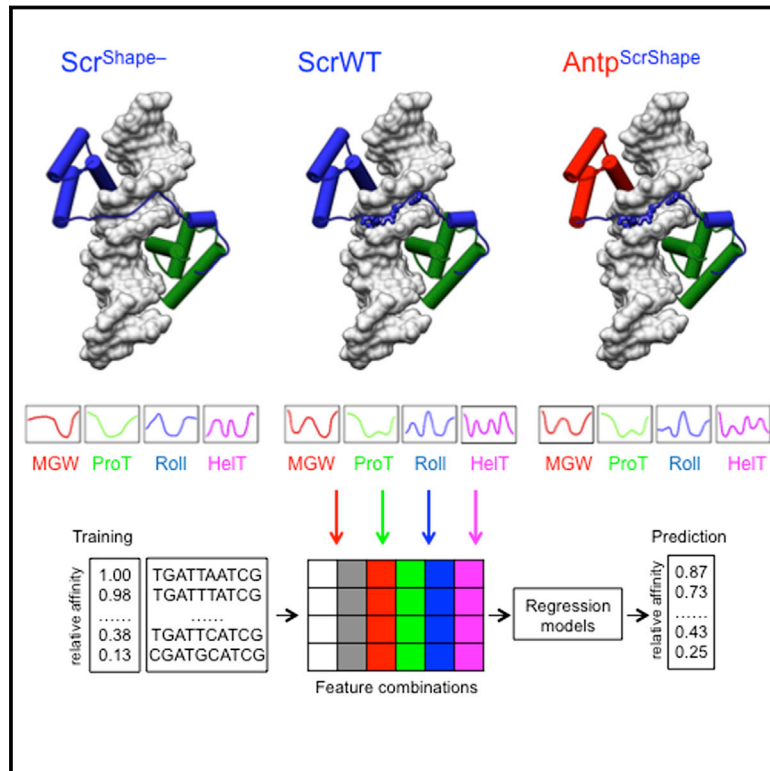
(G) Filipin staining of the cultured fibroblasts from human patients with indicated diseases. Scale bar, 25  $\mu$ m.



- Forss-Petter, S., Werner, H., Berger, J., Lassmann, H., Molzer, B., Schwab, M.H., Bernheimer, H., Zimmermann, F., and Nave, K.A. (1997). Targeted inactivation of the X-linked adrenoleukodystrophy gene in mice. *J. Neurosci. Res.* **50**, 829–843.
- Fukuda, M., and Mikoshiba, K. (2001). Mechanism of the calcium-dependent multimerization of synaptotagmin VII mediated by its first and second C2 domains. *J. Biol. Chem.* **276**, 27670–27676.
- Goldstein, J.L., DeBose-Boyd, R.A., and Brown, M.S. (2006). Protein sensors for membrane sterols. *Cell* **124**, 35–46.
- Hardeman, D., Versantvoort, C., van den Brink, J.M., and van den Bosch, H. (1990). Studies on peroxisomal membranes. *Biochim. Biophys. Acta* **1027**, 149–154.
- Ishibashi, S., Schwarz, M., Frykman, P.K., Herz, J., and Russell, D.W. (1996). Disruption of cholesterol 7 $\alpha$ -hydroxylase gene in mice. I. Postnatal lethality reversed by bile acid and vitamin supplementation. *J. Biol. Chem.* **271**, 18017–18023.
- Jeynov, B., Lay, D., Schmidt, F., Tahirovic, S., and Just, W.W. (2006). Phosphoinositide synthesis and degradation in isolated rat liver peroxisomes. *FEBS Lett.* **580**, 5917–5924.
- Kapitein, L.C., Schlager, M.A., Kuijpers, M., Wulf, P.S., van Spronsen, M., MacKintosh, F.C., and Hoogenraad, C.C. (2010). Mixed microtubules steer dynein-driven cargo transport into dendrites. *Curr. Biol.* **20**, 290–299.
- Kwon, H.J., Abi-Mosleh, L., Wang, M.L., Deisenhofer, J., Goldstein, J.L., Brown, M.S., and Infante, R.E. (2009). Structure of N-terminal domain of NPC1 reveals distinct subdomains for binding and transfer of cholesterol. *Cell* **137**, 1213–1224.
- Lange, Y., Ye, J., Rigney, M., and Steck, T.L. (1999). Regulation of endoplasmic reticulum cholesterol by plasma membrane cholesterol. *J. Lipid Res.* **40**, 2264–2270.
- Liscum, L., and Faust, J.R. (1989). The intracellular transport of low density lipoprotein-derived cholesterol is inhibited in Chinese hamster ovary cells cultured with 3- $\beta$ -[2-(diethylamino)ethoxy]androst-5-en-17-one. *J. Biol. Chem.* **264**, 11796–11806.
- Liscum, L., and Munn, N.J. (1999). Intracellular cholesterol transport. *Biochim. Biophys. Acta* **1438**, 19–37.
- Liu, B., Ramirez, C.M., Miller, A.M., Repa, J.J., Turley, S.D., and Dietschy, J.M. (2010). Cyclodextrin overcomes the transport defect in nearly every organ of NPC1 mice leading to excretion of sequestered cholesterol as bile acid. *J. Lipid Res.* **51**, 933–944.
- Matsumoto, N., Tamura, S., and Fujiki, Y. (2003). The pathogenic peroxin Pex26p recruits the Pex1p-Pex6p AAA ATPase complexes to peroxisomes. *Nat. Cell Biol.* **5**, 454–460.
- Moser, H.W., Raymond, G.V., and Dubey, P. (2005). Adrenoleukodystrophy: new approaches to a neurodegenerative disease. *J. Am. Med. Assoc.* **294**, 3131–3134.
- Prieto Tenreiro, A., Penacho Lazaro, M.A., Andres Celda, R., Fernandez Fernandez, M., Gonzalez Mateo, C., and Diez Hernandez, A. (2013). Dietary treatment for X-linked adrenoleukodystrophy: is “Lorenzo’s oil” useful? *Endocrinología y nutrición* **60**, 37–39.
- Rosenbaum, A.I., Zhang, G., Warren, J.D., and Maxfield, F.R. (2010). Endocytosis of beta-cyclodextrins is responsible for cholesterol reduction in Niemann-Pick type C mutant cells. *Proc. Natl. Acad. Sci. USA* **107**, 5477–5482.
- Sleat, D.E., Wiseman, J.A., El-Banna, M., Price, S.M., Verot, L., Shen, M.M., Tint, G.S., Vanier, M.T., Walkley, S.U., and Lobel, P. (2004). Genetic evidence for nonredundant functional cooperativity between NPC1 and NPC2 in lipid transport. *Proc. Natl. Acad. Sci. USA* **101**, 5886–5891.
- Troffer-Charlier, N., Doerflinger, N., Metzger, E., Fouquet, F., Mandel, J.L., and Aubourg, P. (1998). Mirror expression of adrenoleukodystrophy and adrenoleukodystrophy related genes in mouse tissues and human cell lines. *Eur. J. Cell Biol.* **75**, 254–264.
- Urano, Y., Watanabe, H., Murphy, S.R., Shibuya, Y., Geng, Y., Peden, A.A., Chang, C.C., and Chang, T.Y. (2008). Transport of LDL-derived cholesterol from the NPC1 compartment to the ER involves the trans-Golgi network and the SNARE protein complex. *Proc. Natl. Acad. Sci. USA* **105**, 16513–16518.
- van der Zand, A., and Tabak, H.F. (2013). Peroxisomes: offshoots of the ER. *Curr. Opin. Cell Biol.* **25**, 449–454.
- Vance, J.E., and Vance, D.E. (1990). Lipoprotein assembly and secretion by hepatocytes. *Annu. Rev. Nutr.* **10**, 337–356.
- Weber, T., Zemelman, B.V., McNew, J.A., Westermann, B., Gmachl, M., Parlati, F., Söllner, T.H., and Rothman, J.E. (1998). SNAREpins: minimal machinery for membrane fusion. *Cell* **92**, 759–772.
- Woudenberg, J., Rembacz, K.P., Hoekstra, M., Pellicoro, A., van den Heuvel, F.A., Heegsma, J., van Ijendoorn, S.C., Holzinger, A., Imanaka, T., Moshage, H., and Faber, K.N. (2010). Lipid rafts are essential for peroxisome biogenesis in HepG2 cells. *Hepatology* **52**, 623–633.

# Deconvolving the Recognition of DNA Shape from Sequence

## Graphical Abstract



## Authors

Namiko Abe, Iris Dror, ..., Remo Rohs, Richard S. Mann

## Correspondence

rohs@usc.edu (R.R.),  
rsm10@columbia.edu (R.S.M.)

## In Brief

DNA structural features play an independent and direct role in binding specificity by Hox proteins, and knowledge of these features facilitates the de novo prediction of DNA binding specificities.

## Highlights

- DNA shape-recognizing residues play a direct role in Hox-DNA binding specificity
- These residues are sufficient to swap the specificity of one Hox protein to another
- Accuracy of binding specificity predictions improves by including DNA shape features
- Machine learning reveals positions in the binding site where DNA shape is recognized

## Accession Numbers

GSE65073



# Deconvolving the Recognition of DNA Shape from Sequence

Namiko Abe,<sup>1,2</sup> Iris Dror,<sup>3,7</sup> Lin Yang,<sup>3</sup> Matthew Slattery,<sup>8</sup> Tianyin Zhou,<sup>3</sup> Harmen J. Bussemaker,<sup>9</sup> Remo Rohs,<sup>3,4,5,6,\*</sup> and Richard S. Mann<sup>1,2,\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biophysics

<sup>2</sup>Department of Systems Biology

Columbia University, New York, NY 10032, USA

<sup>3</sup>Molecular and Computational Biology Program, Department of Biological Sciences

<sup>4</sup>Department of Chemistry

<sup>5</sup>Department of Physics and Astronomy

<sup>6</sup>Department of Computer Science

University of Southern California, Los Angeles, CA 90089, USA

<sup>7</sup>Department of Biology, Technion – Israel Institute of Technology, Haifa 32000, Israel

<sup>8</sup>Department of Biomedical Sciences, University of Minnesota Medical School, Duluth, MN 55812, USA

<sup>9</sup>Department of Biological Sciences, Columbia University, New York, NY 10032, USA

\*Correspondence: [rohs@usc.edu](mailto:rohs@usc.edu) (R.R.), [rsm10@columbia.edu](mailto:rsm10@columbia.edu) (R.S.M.)

<http://dx.doi.org/10.1016/j.cell.2015.02.008>

## SUMMARY

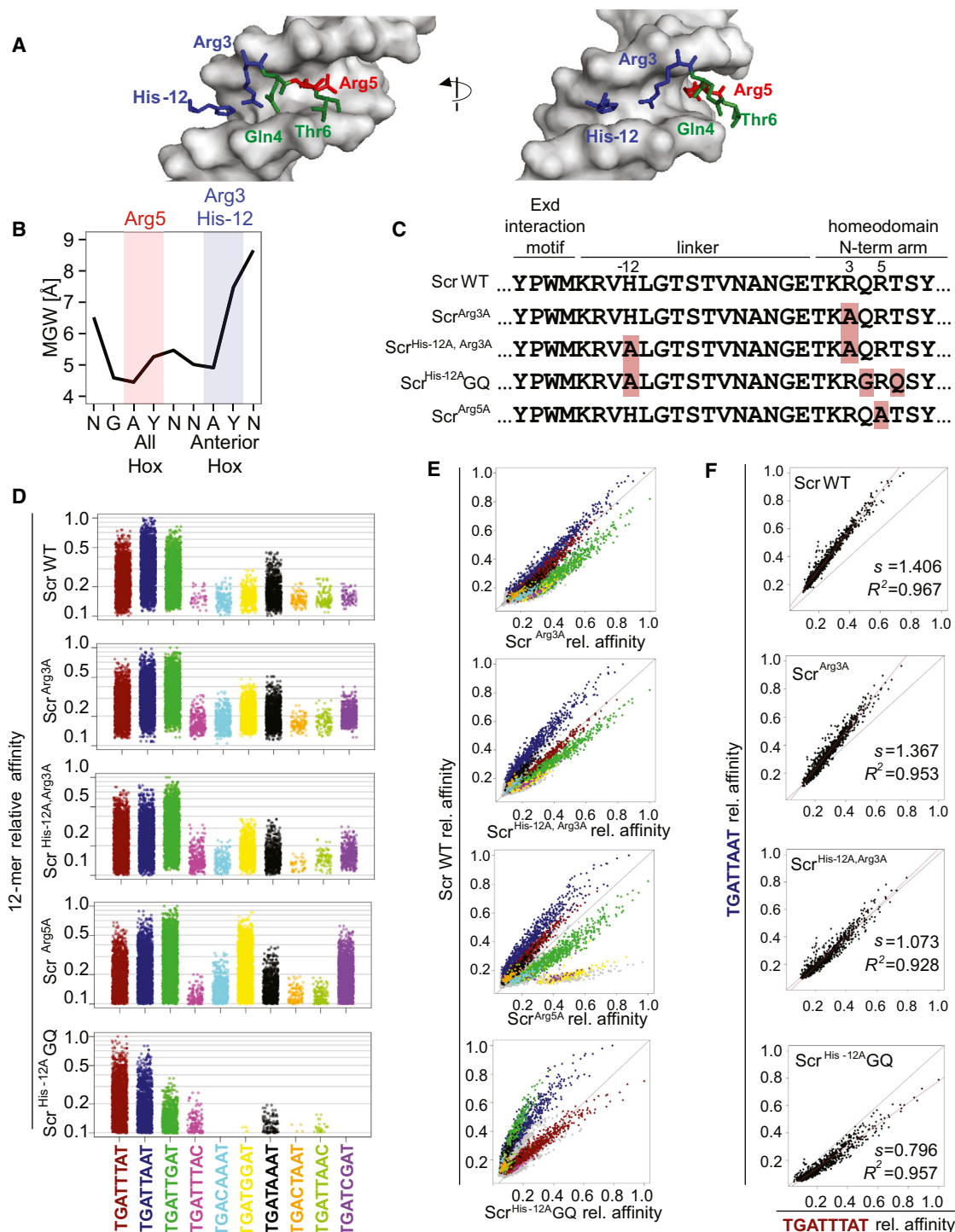
Protein-DNA binding is mediated by the recognition of the chemical signatures of the DNA bases and the 3D shape of the DNA molecule. Because DNA shape is a consequence of sequence, it is difficult to dissociate these modes of recognition. Here, we tease them apart in the context of Hox-DNA binding by mutating residues that, in a co-crystal structure, only recognize DNA shape. Complexes made with these mutants lose the preference to bind sequences with specific DNA shape features. Introducing shape-recognizing residues from one Hox protein to another swapped binding specificities in vitro and gene regulation in vivo. Statistical machine learning revealed that the accuracy of binding specificity predictions improves by adding shape features to a model that only depends on sequence, and feature selection identified shape features important for recognition. Thus, shape readout is a direct and independent component of binding site selection by Hox proteins.

## INTRODUCTION

Precise control of gene expression relies on the ability of transcription factors to recognize specific DNA binding sites. Two distinct modes of protein-DNA recognition have been described: base readout, the formation of hydrogen bonds or hydrophobic contacts with functional groups of the DNA bases, primarily in the major groove (Seeman et al., 1976), and shape readout, the recognition of the 3D structure of the DNA double helix (Rohs et al., 2009a). The importance of shape readout has been inferred from crystal structures of protein-DNA complexes (Joshi et al., 2007; Kitayner et al., 2010; Meijnsing et al., 2009;

Rohs et al., 2009b) and from structural features of DNAs selected by DNA-binding proteins in high-throughput binding assays (Dror et al., 2014; Gordán et al., 2013; Lazarovici et al., 2013; Slattery et al., 2011; Yang et al., 2014). However, as DNA shape is a function of the nucleotide sequence, it is difficult to tease apart whether a DNA binding protein favors a particular binding site because it recognizes its nucleotide sequence or, alternatively, structural features of the DNA molecule. Thus, whether DNA shape is a direct determinant of protein-DNA recognition remains an open question. In addition to being a potentially important mode of DNA recognition, if DNA binding proteins directly use shape readout then incorporating DNA structural information should significantly improve models for predicting DNA binding specificity, which remains challenging with existing methods (Slattery et al., 2014; Weirauch et al., 2013).

We previously described a role for DNA shape in the recognition of specific binding sites by the Hox family of transcription factors, which in vertebrates and *Drosophila* specify the unique characteristics of embryonic segments along the anterior-posterior axis (Joshi et al., 2007; Mann et al., 2009; Slattery et al., 2011). Using in vitro selection combined with deep sequencing (SELEX-seq), which examines millions of sequences in an unbiased manner, we found that while Hox proteins bind highly similar sequences as monomers, heterodimerization with the cofactor Extradenticle (Exd) uncovers latent DNA binding specificities (Slattery et al., 2011). High-throughput DNA shape predictions (Zhou et al., 2013) for sequences selected by each Exd-Hox complex (containing the motif NGAYNNAY) revealed that anterior and posterior Hox proteins prefer sequences with distinct minor groove (MG) topographies. Whereas all Exd-Hox complexes preferred sequences with a narrow MG near the AY of the Exd half-site (NGAY), only anterior Hox proteins (Lab, Pbx, Dfd, and Scr) selected for sequences containing an additional minimum in MG width at the AY of the Hox half-site (NNAY) (Figures 1A and S1) (Slattery et al., 2011). However, this study, as well as analyses of other protein-DNA complexes (Gordán et al., 2013; Yang et al., 2014), did not rule out the



**Figure 1. Scr's Narrow-MG Recognizing Residues Are Required for Binding Specificity**

(A) Two views of the Exd-Scr heterodimer bound to the Scr-specific target *fkh250* (Protein Data Bank [PDB] ID 2R5Z) (Joshi et al., 2007).

(B) Plot of MG width derived from the Exd-Scr co-crystal structure showing that Arg5 (red) inserts into the MG width minimum at the Exd half-site (NGAY) while Arg3 and His-12 (blue) insert into the MG width minimum at the Hox half-site (NNAY).

(C) Amino acid sequences of Scr variants. Numbering is relative to the first residue in the homeodomain. Only sequences from the Exd-interaction motif YPWM through the homeodomain N-terminal arm are shown. Red highlights mutated residues.

(D) 12-mer relative affinities of binding sites selected by each Scr variant in complex with Exd are color-coded according to the ten most frequently observed Exd-Hox binding sites.

(legend continued on next page)



possibility that these shape preferences were merely a secondary consequence of base readout preferences.

A key prediction of the shape-recognition model is that if the residues that recognize a distinct structural feature of the DNA, for example a local minimum in MG width, are mutated then the transcription factor should no longer prefer to bind DNA sequences containing that feature. Alternatively, if the structural feature were merely a byproduct of the DNA sequences selected by a base readout mechanism in the major groove, the binding sites preferred by the mutant factor would still contain that feature. Here, we tested this prediction using the anterior Hox protein Scr, which binds DNA with Exd to regulate Scr-specific target genes during *Drosophila* embryogenesis (Ryoo and Mann, 1999). In a co-crystal structure of the Exd-Scr heterodimer bound to an Scr-specific target site, *fkh250* (AGATTAAT), both shape readout and base readout mechanisms were evident (Joshi et al., 2007). In agreement with the SELEX-seq data, the *fkh250* binding site contained two MG width minima, one recognized by Scr residues His-12 and Arg3, and the second recognized by Scr residue Arg5 (Figures 1A and 1B) (Joshi et al., 2007). As these residues did not form hydrogen bonds with bases, the implication is that they use shape readout, and not base readout, as their sole mode of DNA recognition.

To test if Hox proteins directly use shape readout, we characterized the properties of mutant proteins that, based on the Exd-Scr co-crystal structures, are predicted to either lose or gain the ability to read specific MG topographies. When MG-inserting residues of Scr were mutated to alanines, thus impairing its ability to use shape readout, the mutant proteins no longer preferred sequences containing these MG width minima. Conversely, when MG recognizing residues from Scr were transferred to a Hox protein that normally does not select for this structural feature, the proteins selected binding sites with two MG minima in vitro and gained the ability to activate an Scr-specific target gene in vivo. Finally, we show that taking DNA shape features into consideration significantly improved the ability to predict Exd-Hox binding site specificities compared to models that only depend on DNA sequence. Together, these findings demonstrate that transcription factors directly use shape readout for protein-DNA recognition, and in silico prediction of DNA binding specificities will benefit by taking DNA structural features into consideration.

## RESULTS

### Mutants that Interfere with Scr's Ability to Read MG Shape

In an initial set of experiments to tease apart the contributions of shape readout from base readout, we mutated Scr residues

His-12, Arg3, and Arg5, which, in a co-crystal structure, only use shape readout as their mode of recognition (Joshi et al., 2007) (Figures 1A and 1B). We generated a series of mutant proteins that change these residues to alanines and, consequently, impair Scr's ability to recognize local MG topographies. We mutated either Arg3 alone (Scr<sup>Arg3A</sup>), His-12 alone (Scr<sup>His-12A</sup>), both His-12 and Arg3 (Scr<sup>His-12A, Arg3A</sup>), or Arg5 alone (Scr<sup>Arg5A</sup>) and tested the effect of these mutations in complex with Exd on Scr's DNA binding site preferences using SELEX-seq (Figure 1C).

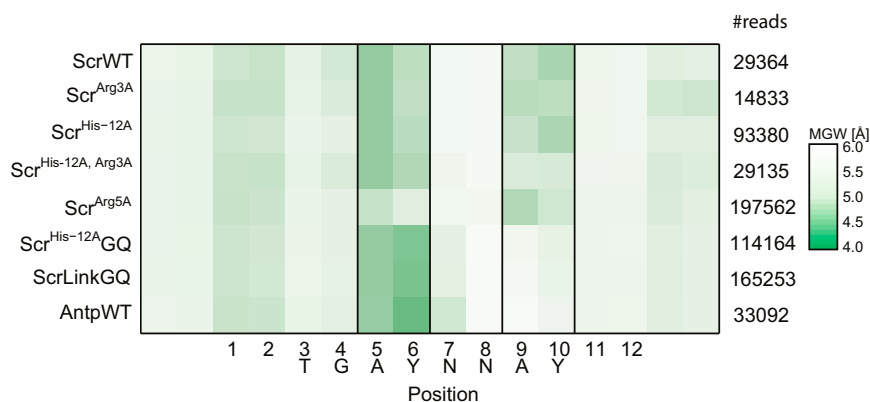
Because the binding site for Exd-Hox complexes is 12 base pairs (Slattery et al., 2011), we generated 12-mer relative affinities for each Scr mutant in complex with Exd using small modifications of our previously described procedure (see Experimental Procedures) (Riley et al., 2014; Slattery et al., 2011), and compared them to the affinities generated by wild-type (WT) Exd-Scr heterodimers. We color-coded the 12-mers based on their core 8-mer (Figure 1D) (Slattery et al., 2011). Compared to Scr WT, all three mutants showed an increased relative preference for the green (TGATTGAT), yellow (TGATGGAT), and purple (TGATCGAT) motifs, and a decrease in the preference for the blue (TGATTAAT) motif (Figures 1D and 1E). Because the blue motif includes the Scr-specific *fkh250* binding site, we directly compared the blue and red (Exd-Hox consensus) motifs by plotting the relative affinities of 12-mer pairs that only differed at the single position that distinguished them from being blue or red (e.g., nnTGATTAATnn with nnTGATTTATnn). Whereas ScrWT showed a preference for blue compared to red motifs over the entire range of affinities, this preference was weakened for Scr<sup>Arg3A</sup> and abolished for Scr<sup>His-12A, Arg3A</sup> (Figure 1F).

Although the above results show that Arg3 and His-12 are required for Scr's binding site preferences, they do not address if this is due to their preference for a specific MG shape. To determine if His-12, Arg3, and Arg5 directly enable the selection of sequences with narrow MGs, we computed the average MG width profile for thousands of 16-mer sequences that were preferentially bound by each Scr variant in our SELEX-seq experiments. We employed DNashape, a high-throughput method for the prediction of the structural features of DNA sequences based on the average conformations of pentamers derived from all-atom Monte Carlo simulations (Zhou et al., 2013). Sequences selected by Scr<sup>His-12A, Arg3A</sup> had an average MG width at A<sub>9</sub> and Y<sub>10</sub> that was significantly wider compared to those selected by ScrWT, without affecting the selection of the MG width minimum at A<sub>5</sub>Y<sub>6</sub> (Figures 2, S2, and S3) ( $p < 2 \times 10^{-16}$ ; Mann-Whitney U test). Sequences selected by the single mutant Scr<sup>Arg3A</sup>, but not Scr<sup>His-12A</sup>, had an intermediate width at A<sub>9</sub>Y<sub>10</sub>, suggesting that His-12 and Arg3 synergistically contribute to MG recognition at the Hox half-site (Figures 2 and S3). Conversely, compared to ScrWT, Scr<sup>Arg5A</sup> selected sequences with a wider

(E) Comparative specificity plots comparing the relative binding affinities of 12-mers selected by Exd-ScrWT (y axis) with each Exd-Scr variant (x axis). Each point represents a unique 12-mer that is color-coded according to the core 8-mer it contains. Gray points represent 12-mers that do not contain any of the ten most common cores. The black line indicates  $y = x$ .

(F) Plots comparing the relative affinities of sequences containing a blue motif (TGATTAAT) (y axis) versus a red motif (TGATTTAT) (x axis) for Exd-ScrWT and Exd-Scr variants. Each point represents the relative affinities of a pair of 12-mers that are identical except for the position that makes it either a blue (nnTGATTAATnn) or a red (nnTGATTTATnn) motif. The black line indicates  $y = x$ , and the red line is a linear regression trend line. The slope of the trend line and coefficient of determination  $R^2$  of the data are indicated.

See also Figures S1, S2, and S3.



**Figure 2. Loss of MG Width Preferences in the Absence of MG-Recognizing Residues**

Heat map of the average MG width at each position of 16-mers selected by each Exd-Hox heterodimer. Dark green represents narrow MG regions whereas white represents wider MG regions. The number of sequences analyzed for each complex is shown on the right. Black lines demarcate where Arg5 inserts into the MG ( $A_5Y_6$ ) and, for ScrWT, where Arg3 and His-12 insert into the MG ( $A_9Y_{10}$ ).

MG specifically in the Exd half-site ( $A_5Y_6$ ), but these sequences retained the minimum at  $A_9Y_{10}$  (Figures 2 and S3). These results provide strong support for the idea that Arg5 directly selects sequences with a MG minimum at  $A_5Y_6$ , while Arg3/His-12 directly select sequences with a MG width minimum at  $A_9Y_{10}$ . Selection of these MG width minima occurs independently, even though they are only separated by two base pairs.

Despite its importance in selecting Scr-specific features of MG topography, Arg3 is present in many Hox homeodomains, including Antennapedia (Antp), that do not select a MG minimum at  $A_9Y_{10}$  (Figures 3A and S1) (Slattery et al., 2011). This observation prompted the question of why Arg3 in Antp and other posterior Hox proteins does not select for a narrow MG at this position. We speculated that the amino acids flanking Arg3 might play a role in binding site selection by correctly positioning this MG-inserting side chain. Indeed, although both Scr and Antp have Arg3 and Arg5, these residues are part of an N-terminal arm motif that differs between these two Hox proteins ( $R_3Q_4R_5T_6$  in Scr and  $R_3G_4R_5Q_6$  in Antp) (Figure 3A). To test whether residues flanking these arginines play a role in Scr binding specificity we characterized an additional mutant, Scr<sup>His-12A</sup>GQ (Figures 1C–1F). In Scr<sup>His-12A</sup>GQ, His-12 is mutated to alanine and the fourth and sixth positions in the Scr homeodomain are changed to that of Antp (Gln<sub>4</sub> to Gly<sub>4</sub> and Thr<sub>6</sub> to Gln<sub>6</sub>) to mimic Antp's  $R_3G_4R_5Q_6$  motif. Strikingly, this mutant failed to select sequences with a minimum at the Hox half site ( $A_9Y_{10}$ ) (Figure 2). An additional mutant, ScrLinkGQ, that, in addition to having Antp's  $R_3G_4R_5Q_6$  motif, has Antp's linker (residues in between the YPWM Exd interaction motif and the homeodomain, Figure S2A) in place of Scr's linker, showed very similar behavior to Scr<sup>His-12A</sup>GQ (Figures 2 and S2). Together, these data suggest that additional residues within and adjacent to the N-terminal arm, which do not make direct contact with the DNA (minor or major groove), play an important role in selecting Hox-specific MG topographies, likely by positioning the MG inserting side chains of Arg3, Arg5, and His-12.

#### Mutants that Transfer Scr's Ability to Read MG Shape to Antp

The above experiments demonstrate that MG-inserting side chains in Scr are necessary for Scr's ability to select sequences with local MG width minima. To test whether MG recognizing res-

idues are sufficient to confer Scr's binding preferences to a different Hox protein, we introduced these residues into Antp, which normally prefers sequences with

wider MG regions at  $A_9Y_{10}$  (Figure 2). We created a series of Antp mutants that contained various combinations of Scr-specific amino acids in two regions, the linker and the N-terminal arm motif  $R_3Q_4R_5T_6$  (Figure 3A). Remarkably, the 12-mer relative affinity profiles of these Antp mutants (in complex with Exd) gradually converged toward that of ScrWT upon the introduction of residues used for MG width recognition (Figures 3B–3D). All of the residues tested—Gln4, Thr6, His-12, and the linker—contributed to the convergence of Antp's binding specificity toward that of Scr, with the most Scr-like mutant, AntpLinkQT, being nearly indistinguishable from ScrWT (Figures 3B and 3C). A direct comparison of the relative affinity for the red motif versus the Scr-preferred blue motif also revealed a gradual shift in preference toward the blue motif (Figure 3D). Thus, Scr-specific amino acids from its linker and N-terminal arm are sufficient to confer Scr's binding specificity on another Hox protein.

To determine if these Antp mutants also share Scr's MG shape preferences, we used DNASHape to predict the MG widths of 16-mers selected by these proteins. In general, the average MG width at  $A_9Y_{10}$  of the sequences selected by the Antp mutant series became narrower, toward that of Scr, upon the introduction of Scr-specific residues (Figure 4A), where, with the exception of AntpHQ, each successive mutant selected sequences with a statistically significant narrowing of the average MG at these positions (Figure S4A). On average, the differences in MG widths at these positions were larger for high-affinity sequences than for low-affinity sequences (Figure S4B). Taken together, these results suggest that Scr residues Gln4, Thr6, His-12, and linker all contribute to the recognition of DNA shape. Moreover, these residues are sufficient to confer the shape preferences of Scr when inserted into another Hox protein.

As an alternative way to analyze these data, we compared the binding specificities of each Antp variant geometrically by calculating the Euclidean distances between the MG width profiles of sequences selected by each variant with the average MG width of those selected by Exd-AntpWT and Exd-ScrWT, respectively. The resulting density plots showed two occupancy peaks, one representing sequences that are more similar to those selected by AntpWT, and a second representing sequences that are more similar to those selected by ScrWT (Figure 4B). With the exception of AntpHQ, each successive Antp variant showed a gradual shift toward the ScrWT peak, with AntpLinkQT showing

a nearly complete shift. Thus, key Scr-specific residues in the linker and N-terminal arm were sufficient to convert Antp's shape preferences to those of Scr.

### Antp Variants that Mimic Scr's DNA Shape Preferences Activate an Scr-Specific Target In Vivo

The above results demonstrate that shape readout, mediated by a limited number of Hox residues, is an essential component of DNA recognition by Exd-Hox heterodimers in vitro. But how relevant is this readout mechanism in vivo? To answer this question we examined the ability of these Antp variants to activate *fkh250-lacZ*, an Scr-specific reporter gene that contains a binding site (AGATTAAT) with two MG width minima (Joshi et al., 2007). In otherwise wild-type embryos, *fkh250-lacZ* expression was limited to Scr-expressing cells in parasegment 2 (PS2) (Ryoo and Mann, 1999) (Figure 5A). Ectopic expression of ScrWT using the *prd-Gal4* driver activated *fkh250-lacZ* in segments outside PS2 (Figure 5B), and His-12 and Arg3 of Scr are required for this activation (Joshi et al., 2007). In contrast to Scr, ectopic expression of AntpWT did not activate *fkh250-lacZ* (Figure 5C). However, ectopic expression of AntpHQT resulted in modest *fkh250-lacZ* activation (Figure 5D), while ectopic expression of AntpLinkQT, the Antp mutant whose binding specificity most closely resembled Scr in vitro (Figures 3 and 4), resulted in strong activation of *fkh250-lacZ* (Figure 5E). Thus, Antp mutants that prefer to bind sequences with two MG width minima in vitro, the normal topography of an Scr-specific binding site, also have the ability to activate an Scr-specific target gene in vivo.

### DNA Shape Features Improve Accuracy of Binding Specificity Predictions

If shape readout is a direct and independent determinant of Hox-DNA binding specificity, we speculated that shape features of the target DNA could be used to improve quantitative predictions of relative binding affinities. To test this notion, we trained an L2-regularized multiple linear regression (MLR) model (Yang et al., 2014) for each of the mutants and WT Hox proteins. We used 10-fold cross validation in order to train and determine the accuracy of a given model, quantified as the coefficient of determination  $R^2$ . These MLR-derived  $R^2$ s are robust as they are highly correlated with  $R^2$ s derived using an alternative machine learning approach, support vector regression ( $\epsilon$ -SVR) with a linear kernel (Figure S5; see Experimental Procedures for details) (Gordân et al., 2013; Zhou et al., 2015).

Using MLR, addition of MG width to a model based only on nucleotide sequence resulted in a modest improvement in  $R^2$  of on average 12% (Figures 6A and S6A). Like MG width, adding three other shape features one at a time, Roll, propeller twist (ProT), and helix twist (HelT), also led to a modest improvement in accuracy (Figure S6A). Inclusion of all four DNA shape features in combination further increased prediction accuracy (Figures 6B and S6A). The improvement in binding affinity prediction accuracy, on average 26% when incorporating all four shape features, yielded the largest effect with high significance ( $p = 6 \times 10^{-5}$ ; Mann-Whitney U test). The addition of any combination of three shape features led to an intermediate increase in prediction accuracy, in some cases similar to that after addition of all

four shape features (Figure S6A). These results suggest that all four features contribute to Exd-Hox-DNA target selection in a non-additive manner, consistent with the interdependency of these features (Olson et al., 1998). Thus, including DNA shape features in addition to MG width improves binding site predictions over models based only on nucleotide sequence.

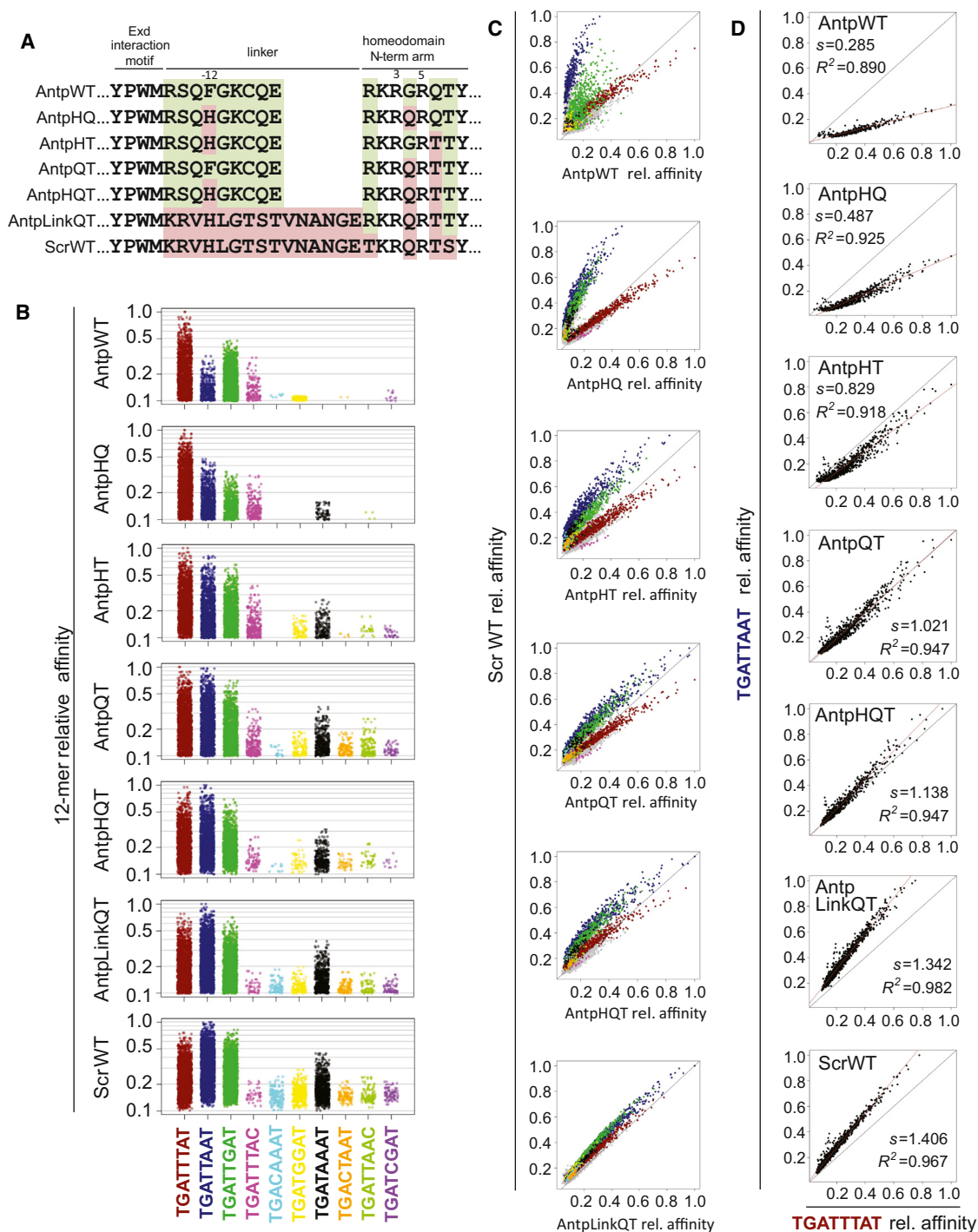
For comparison, we also assessed the benefit of adding shape features for the prediction of Hox monomer specificities. Interestingly, in this case the improvement in  $R^2$  was, on average, only 6.4%, suggesting a larger role for DNA shape in conferring heterodimer specificity than monomer specificity (Figure S6B).

### DNA Shape Contributes to Binding Specificities in a Position-Specific Manner

Next, we hypothesized that if shape features contribute to an improvement in binding specificity prediction, then it might be possible to localize this effect within the binding site. We trained models using the sequence of the entire binding site augmented by all four shape features at individual positions one at a time, resulting in a set of models that tested the contribution of shape at each position of the binding site. We compared these models to a sequence-only model and calculated a  $\Delta R^2$ . This analysis highlighted the importance of DNA shape for predicting Exd-Hox binding specificities in the core, but not the flanks, of the binding site (Figure 6C).

To analyze the role of DNA shape in a complementary manner, we trained shape-only models using the four shape features at all nucleotide positions, leaving out this information one position at a time, resulting in a set of models that assessed the relative importance of DNA shape at each position of the binding site.  $\Delta R^2$ s were calculated relative to a model that included the four shape features at all positions. In this analysis, prediction accuracy was expected to decrease most when shape features were removed from the model at positions that were important for shape readout. Interestingly, we detected the greatest effect at the A<sub>9</sub> position of the Hox half-site, followed by slightly weaker effects at the adjacent Y<sub>10</sub> position and the G<sub>4</sub> position of the Exd half-site (Figure 6D). Eliminating shape features from the remaining positions had a smaller impact on the ability to predict binding specificities.

Within each SELEX-seq data set, the sequences were most variable at the N<sub>8</sub> position, raising the possibility that the success of these models might be driven in large part by this position. To test this idea and better assess the role of DNA shape throughout the binding site we trained additional models in which we removed sequence information at the N<sub>8</sub> position ("sequence-N<sub>8</sub> model"). Leaving out sequence information at the N<sub>8</sub> position did not significantly affect the accuracy of a sequence+shape model, suggesting that sequence information at N<sub>8</sub> is not essential for its performance (Figure S7A). When MG width information was added to the sequence-N<sub>8</sub> model, the ability to predict binding specificities was greatly enhanced compared to the same model without MG width information (Figures S7B and S7C). These results argue that MG width information is more important than sequence at positions with a degenerate sequence signal, such as at N<sub>8</sub>, where direct readout is not playing a role. The removal of the confounding sequence information at this position uncovered MG width as an independent specificity determinant.



**Figure 3. Introducing Scr's MG Width-Recognizing Residues into Antp Converts Its Binding Specificity to that of Scr**

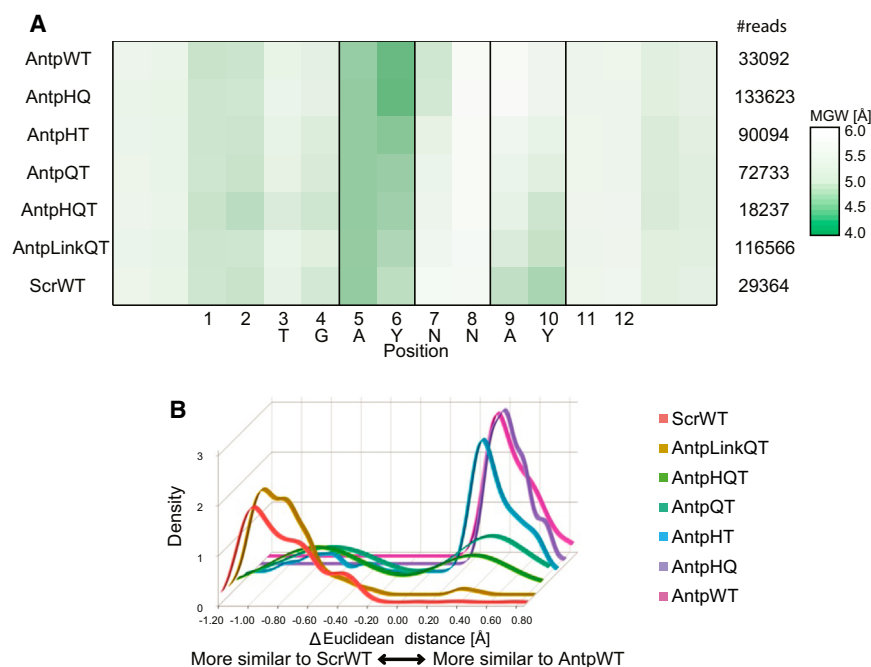
(A) Amino acid sequences (from the Exd interaction motif, YPWM, through the N-terminal arm of the homeodomain) of Antp variants. Green highlights residues specific to AntpWT, and red highlights residues specific to ScrWT. Non-highlighted residues are common between the two Hox proteins. Numbering is relative to the first residue of Scr's homeodomain. The rest of the protein is wild-type in all cases.

(B) 12-mer relative affinities of binding sites selected by each Antp variant in complex with Exd are color-coded according to the ten most commonly observed Exd-Hox motifs. AntpWT and ScrWT are included to show the progression of the binding preferences from AntpWT toward ScrWT.

(C) Comparative specificity plots of the relative affinity of sequences selected by Exd-ScrWT (y axis) and each Exd-Antp mutant (x axis). Each point represents a unique 12-mer that is color-coded according to the core 8-mer it contains. Gray points represent 12-mers that do not contain any of the ten most common cores. The black line indicates  $y = x$ .

(legend continued on next page)





**Figure 4. Shape Readout Properties of Antp Variants with Scr-Specific Residues**

(A) Heat map of the average MG width at each position of all statistically significant 16-mers selected by each Exd-Hox complex. Dark green represents narrow MG regions whereas white represents wider MG regions. The number of sequences analyzed for each protein is shown on the right. Black lines demarcate where Arg5 inserts into the MG ( $A_5Y_6$ ) and, for Scr, where Arg3 and His-12 insert into the MG ( $A_9Y_{10}$ ).

(B) Histogram representing the distribution of MG width similarities for each of the sequences selected by each Antp variant in comparison to those selected by ScrWT and AntpWT. The y axis represents the density of 16-mers at different  $\Delta$ (Euclidean distance) scores (x axis). Sequences more similar to those selected by ScrWT receive a negative score, and sequences more similar to those selected by AntpWT receive a positive score. See also Figure S4.

When all four shape features were added to the sequence- $N_8$  model at single positions one at a time the contribution of DNA shape within the core motif was very apparent (Figure 7A), and significantly stronger than when the starting model included sequence information at the  $N_8$  position (compare with Figure 6C). If instead of all four DNA shape parameters only MG width was added position by position to the sequence- $N_8$  model, the average improvement in  $R^2$ , while smaller, was most apparent at or adjacent to  $Y_6N_7$  and  $A_9Y_{10}$  (Figure 7B). Thus, although DNA shape is generally important within the entire core of the binding site, the contribution of MG width is strongest at the two AY regions, precisely where local minima in MG width were observed in the Exd-Hox X-ray structures (Joshi et al., 2007) and SELEX-seq data (Figures 2 and 4).

Taken together, quantitative predictions based on regression models indicated that shape features become important where sequence information is not well defined, more likely at positions that are not involved in base readout. In these cases, shape features contain more information than sequence alone, and removing the signal from sequence enables the quantitative modeling of the role of shape features on binding specificity.

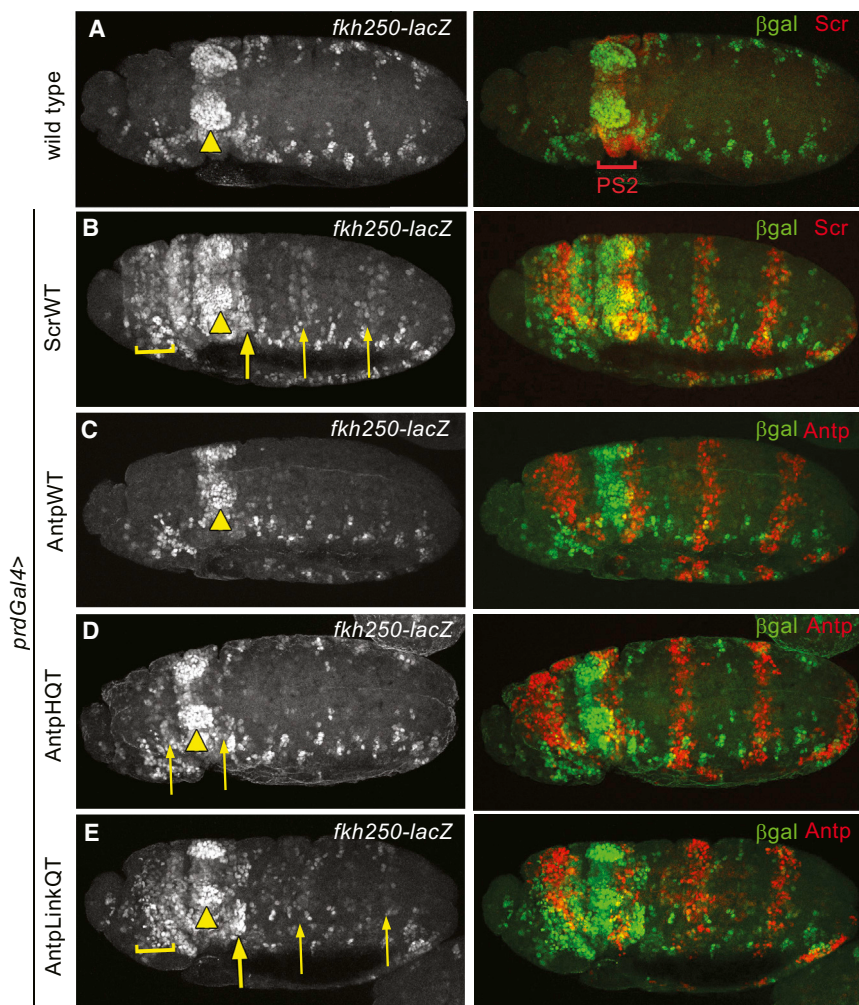
### DNA Shape Features Discriminate Anterior from Posterior Hox Binding Specificities

To understand to what extent shape features can help distinguish Exd-ScrWT from Exd-AntpWT binding specificities, we assigned a value of +1 to the top 50% of sequences selected by Exd-ScrWT and -1 to the top 50% of sequences selected by Exd-AntpWT

(see Experimental Procedures for details). We then used sequence- and shape-based models to evaluate the discriminative power of the selected features. Using L2-regularized MLR and 10-fold cross validation, we calculated the area under the receiver-operating characteristic curve (AUC) as a criterion for a model to discriminate ScrWT-like from AntpWT-like binding specificities. We found that MG width alone, without using sequence or additional shape features, discriminates between the binding specificities of both Exd-Hox complexes with high accuracy (Figure S7D). Thus, MG width does not merely refine binding specificity but is a powerful descriptor on its own, at least for discriminating between these two Exd-Hox complexes. Classification models using other shape parameters performed similarly well (Figure S7D), indicating that a classification between two states is less sensitive than quantitative prediction of binding strength using regression models. Further, these results suggest that the qualitative differences that are apparent in the MG width heat maps (Figures 2 and 4A) reflect a quantitative difference in anterior and posterior Hox specificities.

Next, we asked which positions in the binding site had the highest impact on this classification. To answer this question, we calculated the Pearson correlation between the class labels +1 and -1 for Exd-ScrWT and Exd-AntpWT, respectively, and MG width at each position (see Experimental Procedures for details). Several positions showed strong, either positive or negative, correlations that enabled the classification into ScrWT-like and AntpWT-like binding specificities (Figure 7C). Two regions showing a negative Pearson correlation aligned with the two MG width minima observed in the Exd-Scr co-crystal structure, and a region of positive Pearson correlation marked

(D) Plots comparing the relative affinities of sequences containing a blue motif (TGATTAAT) (y axis) versus a red motif (TGATTAT) (x axis) for ScrWT, AntpWT and Antp variants. Each point represents the relative affinities of a pair of 12-mers that are identical except for the position that makes it either a blue (TGATTAAT) or a red (TGATTAT) motif. The black line indicates  $y = x$ , and the red line is a linear regression trend line. The slope of the trend line and coefficient of determination  $R^2$  of the data are indicated.



**Figure 5. Scr's MG Width Readout Residues Confer the Ability to Activate an Scr-Specific Target In Vivo when Incorporated into Antp**

(A) In wild-type embryos *fkh250-lacZ* is activated only in parasegment 2 (PS2), where endogenous Scr is expressed (arrowhead). In this and all panels, anterior is to the left.

(B) Ectopic expression of ScrWT using *prd-Gal4* (visualized with red stripes of ectopic expression in the panel on the right) activates *fkh250-lacZ* anterior and posterior to PS2. Activation is strongest anterior to PS2 (bracket) and immediately posterior to PS2 (thick arrow), with weaker activation in abdominal segments (thin arrows).

(C) Ectopic expression of wild-type Antp does not activate *fkh250-lacZ*.

(D) Ectopic expression of AntpHQT leads to weak ectopic *fkh250-lacZ* expression anterior and posterior to PS2 (thin arrows).

(E) Ectopic expression of AntpLinkQT leads to activation both anterior and posterior to PS2. Activation is strongest anterior to PS2 (bracket) and immediately posterior to PS2 (thick arrow), with weaker activation in abdominal segments (thin arrows).

the region between these minima. This observation confirms that the core region is important for the differences in binding specificity between paralogous Hox factors. Interestingly, not only is the AY region of the Hox half-site important, but the shape of the entire core, presumably due to the influence of all core positions on the shape of this region.

Finally, we used classification models to predict whether the DNA shape mutants defined in Figures 1, 3, and S2 tend to show ScrWT-like or AntpWT-like binding specificities. Here, a sequence was classified as ScrWT-like if the class label was predicted to be >0, and as AntpWT-like if the class label was predicted to be <0. This classification indicated a gradual change in the fraction of sequences selected by any of the mutants assigned as ScrWT- versus AntpWT-preferred sequences (Figure 7D). These data quantitatively confirm the qualitative observations shown above (Figures 2 and 4A) that MG width topography is an important binding specificity signal for Hox proteins.

## DISCUSSION

Despite significant effort in the field, it is still not possible to accurately decipher the regulatory information that is encoded

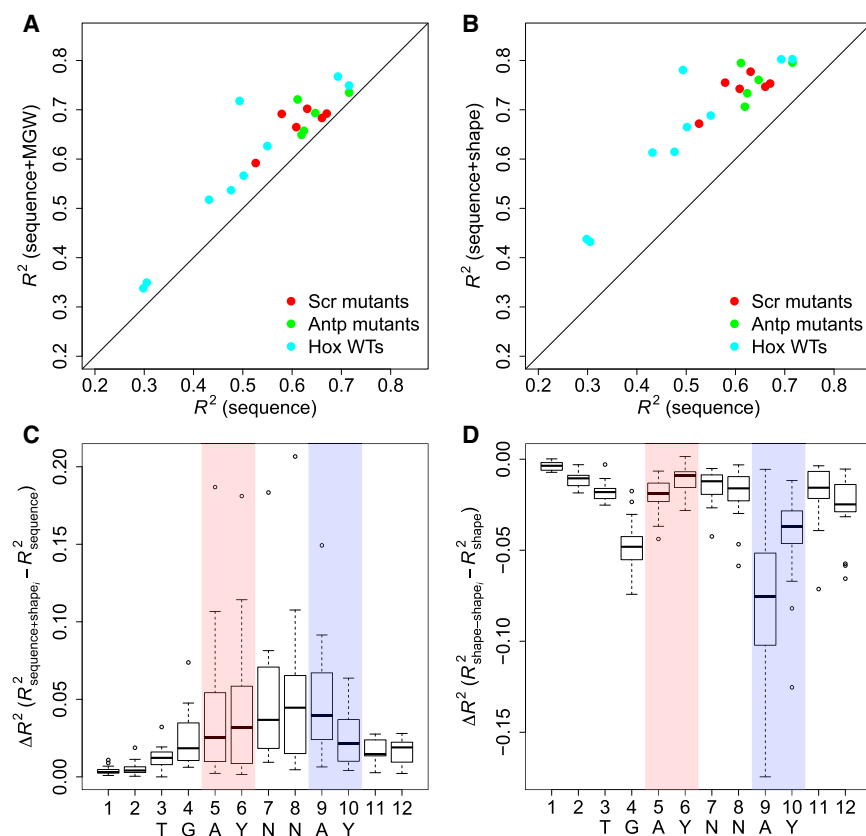
in the DNA sequences of eukaryotic genomes (Slattery et al., 2014). In the work described here, we used a combination of in vitro, in vivo, and computational approaches to show that intrinsic DNA structural characteristics—collectively referred to as DNA shape—are being directly read by DNA binding proteins when they recognize their binding sites.

Thus, analogous to mechanisms in which

DNA base pairs are directly read by proteins via hydrogen bonds, the recognition of DNA shape independently contributes to both binding affinity and specificity. Using this information, we show that including DNA shape features significantly enhances the ability to predict DNA binding specificities and thus will greatly improve models for accurately predicting transcription factor binding in eukaryotic genomes.

## Separable Contributions of DNA Shape and Sequence to Protein-DNA Recognition

Although several previous reports suggested the importance of DNA shape in protein-DNA recognition, all prior work was unable to definitively discriminate between the roles of DNA shape and sequence. Although DNA shape features, such as MG width, were previously found to contribute to binding specificity (Dror et al., 2014; Gordân et al., 2013; Lazarovici et al., 2013; Yang et al., 2014), here the roles of DNA sequence and shape have been separated and analyzed in an unbiased manner. To achieve this, we mutated Scr amino acid side chains that do not make direct base contacts in the major groove, but instead either insert into the MG (His-12, Arg3, Arg5) or indirectly influence these interactions (Gln4, Thr6, linker). The combination of SELEX-seq



**Figure 6. DNA Shape Features Improve Quantitative Predictions of DNA Binding Specificities of Exd-Hox Heterodimers**

(A) Scatter plot representing the coefficient of determination  $R^2$  obtained using a sequence-only model (x axis) compared to a model using sequence and MG width (y axis). Each point represents a different Exd-Hox heterodimer and is color-coded as indicated.

(B) Scatter plot representing the coefficient of determination  $R^2$  obtained using a sequence-only model (x axis) compared to a model using sequence and four DNA shape features (MG width, Roll, ProT and HelT) (y axis). Quantitative measures for the improvement of the prediction accuracy of the logarithm of relative binding affinities using shape-augmented models are provided in Figure S6.

(C) Box plots illustrating the contribution from DNA shape features to model accuracy when shape features were added to a sequence model at each position individually. The effect on the coefficient of determination  $\Delta R^2$  is shown for adding four shape features (MG width, Roll, ProT and HelT) position-by-position to the sequence model. The centerline of the box plots represents the median, the edge of the box the first and third quartile, and the whiskers indicate minimum/maximum values within 1.5 times the interquartile from the box.

(D) Box plots illustrating the contribution from DNA shape features to model accuracy when sequence features were removed. The effect on the coefficient of determination  $\Delta R^2$  is shown for leaving out four shape features (MG width, Roll, ProT, and HelT) position-by-position from a shape-only model that does not contain any sequence information. The box plots are defined in (C).

See also Figures S5 and S6.

with high-throughput DNA shape analysis allowed us to show the effect of these mutations on the selection of DNA binding sites with distinct shape characteristics. Further, not only were these amino acid side chains necessary for conferring the DNA binding preferences of these proteins, they were sufficient to confer this specificity, both in vitro and in vivo, when grafted into a different Hox protein, Antp. These experiments effectively tease apart the contributions of shape readout from base readout. We speculate that the readout of DNA shape may be a general mechanism that transcription factors use to recognize their binding sites. Moreover, for transcription factors that are members of large paralogous families, such as the Hox proteins, DNA shape may be essential for distinguishing between binding sites that are difficult to discriminate based on base readout alone.

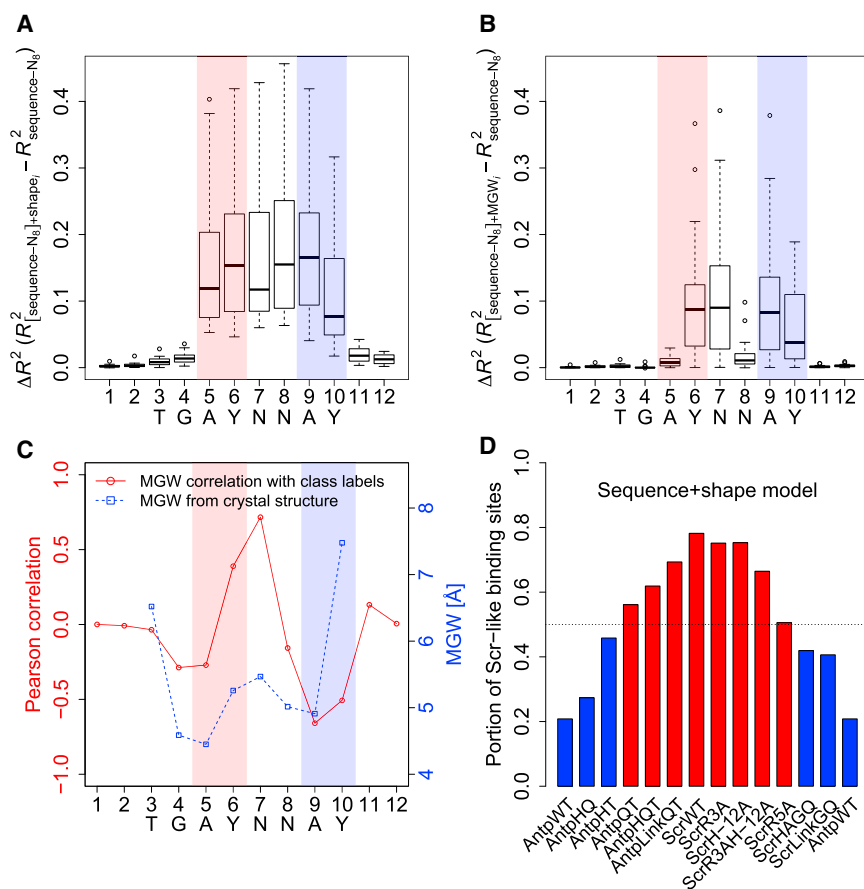
### Statistical Machine Learning Reveals DNA Structure-Based Binding Specificity Signals

To complement and extend the in vitro and in vivo studies, we used statistical machine learning, in this case multiple linear regression (MLR), to computationally analyze the contributions of DNA sequence and shape. Using this approach we were able to (1) quantify the overall contribution of shape features to binding specificity and (2) compute the relative contributions of DNA shape and sequence at individual positions within the bind-

ing site. Extensive experimental work, involving structure determination and mutagenesis, represents the current standard approach for uncovering DNA readout mechanisms of transcription factors. The quantitative modeling introduced here suggests an alternate route for deriving such mechanistic information from high-throughput sequencing data. These methods will therefore likely be valuable when used to predict the DNA binding specificities of other transcription factors and when analyzing their interactions with genomes.

To identify positions in the binding site where shape features contribute substantially to binding specificity, we used a form of feature selection in which we compared models with different feature sets by computing a  $\Delta R^2$  relative to a reference model. We found that the shape features in the core of the Exd-Hox heterodimer binding site were important for paralogous binding specificity. This observation is distinct from previous observations for another family of transcription factors, basic helix-loop-helix (bHLH) factors, where shape features in regions flanking the core binding site play an important role in discriminating binding specificities of related family members in yeast (Gordân et al., 2013) and human (Yang et al., 2014). Further, our feature selection approach indicates that shape features at the AY region of the Hox half-site were the most critical for determining binding specificity. This finding agrees with qualitative





**Figure 7. Models that Deconvolve DNA Sequence and Shape**

(A) Removing sequence features at the  $N_8$  position where sequence is least constrained across the selected sequences from the sequence+shape model further emphasizes the contribution of adding DNA shape to model accuracy. Whereas removing sequence information at this position has essentially no effect on model accuracy (Figure S7A), adding MG width to the sequence- $N_8$  model has a large effect on prediction accuracy (Figure S7B). Based on this finding, the effect on the coefficient of determination  $\Delta R^2$  is shown in box plots for adding four shape features (MG width, Roll, ProT, and HelT) position-by-position to the sequence- $N_8$  model. The centerline of the box plots represents the median, the edge of the box the first and third quartile, and the whiskers indicate minimum/maximum values within 1.5 times the interquartile from the box.

(B) Box plots illustrating the effect on the coefficient of determination  $\Delta R^2$  for adding MG width information position-by-position to the sequence- $N_8$  model emphasize the role of the AY and immediately adjacent positions. The box plots are defined in (A). (C) Pearson correlations (red) between MG width (MGW) and binding site labels (+1 for ScrWT-like versus -1 for AntpWT-like) track with the MGW pattern (blue) observed in the co-crystal structure (Joshi et al., 2007), emphasizing the important role of MGW in the core region of Exd-Hox binding site. (D) A sequence+shape classification model captures the gradual change of binding specificities introduced by mutations of the N-terminal arm and linker sequences with some Exd-Hox mutant heterodimer specificities classified as Scr-like (red) and others as Antp-like (blue). See also Figure S7.

observations in a previous study (Slattery et al., 2011) and in this work (Figures 2 and 4) that shape selections varied most substantially at this position for both wild-type and mutant Hox proteins. While this was previously a qualitative observation, the current study shows the effect quantitatively. The machine learning and feature selection methods reveal that this information will likely provide a powerful approach when analyzing data from high-throughput binding assays for other transcription factors. In particular, it is noteworthy that we were able to derive structural mechanisms used by Hox transcription factors based only on sequence data alone, without solving a 3D structure.

### Broader Implications for Recognition of Genomic Target Sites by Transcription Factors

Based on our findings, we propose that as more high-throughput DNA binding data become available (Hume et al., 2015; Jolma et al., 2013; Zhu et al., 2011), DNA shape parameters should be taken into consideration when analyzing and subsequently scanning genomes for DNA binding site preferences. Further, although different families of transcription factors may use DNA shape in various ways, this information may be used to inform binding site prediction algorithms. As shown here quantitatively, Exd-Hox heterodimers use distinct

structural features in the DNA, such as local regions of narrow MG, to achieve DNA binding specificity. Because MG width minima are distinct structural motifs, we were able to separate their contributions to DNA recognition both biochemically, by mutating amino acids that recognize these motifs, and computationally, by training models that include or exclude specific subsets of DNA features. For other protein families, the contribution of DNA structure may not be as readily separable as it is for Exd-Hox binding. For example, although previous work demonstrated a role for DNA shape in conferring the binding specificity of bHLH proteins, this effect was mediated by sequences flanking the core binding site (E-box), where no known protein-DNA interactions (base or shape readout) occur (Gordán et al., 2013). In this case, the role of DNA shape may be biochemically inseparable from base readout because it is unlikely that a distinct structural motif is formed by the flanking sequences.

Our results have implications for the design of binding site search and de-novo motif discovery methods, which currently most typically rely only on DNA base features (Weirauch et al., 2013). There are some examples where large sets of overlapping DNA structural features, which are highly interdependent from each other and inseparable from sequence, have been integrated in motif search algorithms (Hooghe



et al., 2012; Maienschein-Cline et al., 2012; Meysman et al., 2011). The results described here, however, suggest that for some transcription factor families, distinct structural motifs, which can be defined independently from sequence, such as MG topography, can be directly integrated in genome analysis tools as quantifiable search parameters. The ability to independently define and quantify the role of distinct structural motifs will likely yield more powerful algorithms that may help identify low affinity, high specificity Hox binding sites that are unrecognizable with standard approaches (Crocker et al., 2015). Further, machine learning approaches may also contribute to more accurate models of cooperative transcription factor binding, for example in the interferon- $\beta$  enhanceosome (Chang et al., 2013), or in vivo, where DNA shape has been identified as a predictive feature for transcription factor binding (Barozzi et al., 2014). We further propose that the computational approaches described here will also be valuable for deconvolving and discovering the roles of DNA shape and sequence even for transcription factors such as the bHLH factors where DNA shape cannot be as readily separated biochemically from DNA sequence. The ability to quantitatively assess the distinct roles of DNA sequence and shape will therefore advance our ability to identify bona fide genomic binding sites and the ability to interpret eukaryotic genomes.

## EXPERIMENTAL PROCEDURES

### Oligonucleotides

All oligonucleotides used in this study are listed in Table S1.

### Protein Purification

Scr and Antp mutants were cloned using the QuickChange Site-Directed Mutagenesis Kit (Agilent) using his-tagged ScrWT (Joshi et al., 2007) and his-tagged AntpWT (Jaffe et al., 1997; Noro et al., 2006) as templates. His-tagged proteins were expressed in BL21 cells and purified using Cobalt chromatography. For the SELEX-seq experiments, “Exd” refers to Exd co-purified with the HM domain of Homothorax (Hth) (Noro et al., 2006).

### In Vivo Analysis

All transgenic UAS lines were generated using the  $\phi$ -C31 integration system into the attP2 insertion site. UAS lines were crossed to flies containing *fkh250-lacZ* on the second chromosome and *prd-Gal4* on the third chromosome. Embryos were collected at 25°C and stained using rabbit anti- $\beta$ -galactosidase (Cappell) and either mouse anti-Scr (gift from D. Andrews) or mouse anti-Antp (8C11; DSHB).

### SELEX-Seq

All SELEX experiments were carried out as described (Riley et al., 2014; Slattery et al., 2011). In total, five 16-mer libraries were used for multiplexing (Table S1). Sequencing was performed by Illumina HiSeq 2000/2500. The number of sequences analyzed for each protein is listed in Tables S2 and S3.

### Inferring Relative Binding Affinities

Fifth order Markov models were constructed using Round 0 (R0) sequences to predict the number of 12-, 14-, and 16-mer sequences in each initial library as described (Riley et al., 2014; Slattery et al., 2011). R3 data were used for all Hox variants in order to optimize counts and minimize sampling error. 12-, 14-, and 16-mer relative binding affinities were generated by taking the cubic root of the enrichment ratio (counts in R3 divided by expected counts as predicted using Markov model derived from R0 data).

### High-Throughput DNA Shape Prediction

All sequences selected in R3 of SELEX with a count of at least 25 were aligned based on the TGAYNNAY (Exd-Hox heterodimers) or TAAT (Hox monomers) motifs. Four DNA structural features were derived for these sequences from a high-throughput DNA shape prediction method (Zhou et al., 2013). Euclidean distance was used to compare MG width profiles of sequences selected by Hox mutants to the average MG width at all positions of sequences selected by the Hox WTs. See Extended Experimental Procedures for details.

### Regression Models for Predicting Binding Specificities Quantitatively

To predict the relative binding affinity for sequences bound by the Hox monomers and Exd-Hox heterodimers, we trained L2-regularized multiple linear regression (MLR) models (Yang et al., 2014). A 10-fold cross-validation was performed with an embedded 10-fold cross-validation on the training set to determine the optimal  $\lambda$  parameter. We trained models that (1) encoded the nucleotide sequence of each of the bound sequences as binary features (sequence models), (2) encoded different combinations of the DNA shape features MG width, ProT, Roll, and HelT (shape models), and (3) combined nucleotide sequence and DNA shape features at the corresponding position (sequence+shape models). We calculated the coefficient of determination  $R^2$  between the predicted and experimentally determined logarithm of relative binding affinities using 10-fold cross validation. We used all 14-mer sequences from R3 of the selection with a count of >50, aligned based on the TGAYNNAY core motif for heterodimers, and the logarithm of the relative binding affinity as response variable.  $\Delta R^2$ s were defined as described in the text. See Extended Experimental Procedures for details and access to the source code for DNA shape prediction and feature mapping.

### Classification Models for Distinguishing Binding Specificities

To classify Hox binding specificities, we aligned 14-mers selected by Exd-ScrWT (assigned the label +1) or Exd-AntpWT (assigned the label -1) according to the presence of a single core motif TGAYNNAY. We trained classification models using L2-regularized MLR and used the resulting models to classify the top 50% aligned binding sites preferred by the mutants. The models were evaluated based on this training data using L2-regularized MLR and 10-fold cross-validation, and area under the receiver-operating characteristic curve (AUC) was used as performance measure. See Extended Experimental Procedures for details.

### ACCESSION NUMBERS

The Gene Expression Omnibus (GEO) accession number for the SELEX-seq data sets reported in this paper is GSE65073.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.008>.

### AUTHOR CONTRIBUTIONS

The experiments were conceived by N.A. and R.S.M. The computational approaches were conceived by I.D., L.Y., and R.R. N.A. executed the SELEX-seq experiments with early contributions from M.S. I.D., L.Y., and T.Z. executed the computational analyses. L.Y. designed and implemented the machine learning and feature selection approaches. N.A. and H.J.B. analyzed the SELEX data. N.A., R.S.M., and R.R. wrote the paper.

### ACKNOWLEDGMENTS

We thank Barry Honig, David Stern, and members of the R.R., H.J.B., and R.S.M. laboratories for feedback and comments on this project, and Vince FitzPatrick, Gabriella Martini, Todd Riley, and Roumen Voutev for technical assistance. This work was supported by the NIH (grants R01GM058575 to

R.S.M., F32GM099160 to N.A., R01GM106056 and U01GM103804 to R.R., and R01HG003008 to H.J.B. and R.R.), the USC-TechNion Visiting Fellows Program, and an Alfred P. Sloan Research Fellowship (to R.R.).

Received: August 5, 2014

Revised: December 8, 2014

Accepted: January 26, 2015

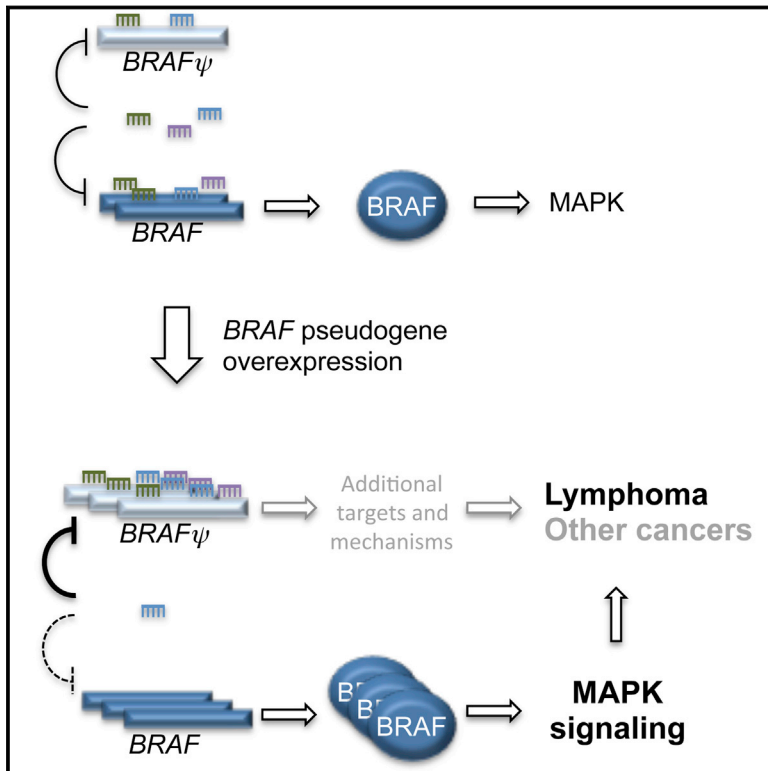
Published: April 2, 2015

## REFERENCES

- Barozzi, I., Simonatto, M., Bonifacio, S., Yang, L., Rohs, R., Ghisletti, S., and Natoli, G. (2014). Coregulation of transcription factor binding and nucleosome occupancy through DNA features of mammalian enhancers. *Mol. Cell* 54, 844–857.
- Chang, Y.P., Xu, M., Machado, A.C., Yu, X.J., Rohs, R., and Chen, X.S. (2013). Mechanism of origin DNA recognition and assembly of an initiator-helicase complex by SV40 large tumor antigen. *Cell Rep.* 3, 1117–1127.
- Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsa-wadi, A., Valenti, P., Plaza, S., Payre, F., et al. (2015). Low affinity binding site clusters confer Hox specificity and regulatory robustness. *Cell* 160, 191–203.
- Dror, I., Zhou, T., Mandel-Gutfreund, Y., and Rohs, R. (2014). Covariation between homeodomain transcription factors and the shape of their DNA binding sites. *Nucleic Acids Res.* 42, 430–441.
- Gordán, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M.L. (2013). Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep.* 3, 1093–1104.
- Hooghe, B., Broos, S., van Roy, F., and De Bleser, P. (2012). A flexible integrative approach based on random forest improves prediction of transcription factor binding sites. *Nucleic Acids Res.* 40, e106.
- Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., and Bulyk, M.L. (2015). UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* 43 (Database issue), D117–D122.
- Jaffe, L., Ryoo, H.D., and Mann, R.S. (1997). A role for phosphorylation by casein kinase II in modulating Antennapedia activity in *Drosophila*. *Genes Dev.* 11, 1327–1340.
- Jolma, A., Yan, J., Whittington, T., Toivonen, J., Nitta, K.R., Rastas, P., Morgunova, E., Enge, M., Taipale, M., Wei, G., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327–339.
- Joshi, R., Passner, J.M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M.A., Jacob, V., Aggarwal, A.K., Honig, B., and Mann, R.S. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131, 530–543.
- Kitayner, M., Rozenberg, H., Rohs, R., Suad, O., Rabinovich, D., Honig, B., and Shakked, Z. (2010). Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.* 17, 423–429.
- Lazarovici, A., Zhou, T., Shafer, A., Dantas Machado, A.C., Riley, T.R., Sandstrom, R., Sabo, P.J., Lu, Y., Rohs, R., Stamatoyannopoulos, J.A., and Bussemaker, H.J. (2013). Probing DNA shape and methylation state on a genomic scale with DNase I. *Proc. Natl. Acad. Sci. USA* 110, 6376–6381.
- Maienschein-Cline, M., Dinner, A.R., Hlavacek, W.S., and Mu, F. (2012). Improved predictions of transcription factor binding sites using physicochemical features of DNA. *Nucleic Acids Res.* 40, e175.
- Mann, R.S., Lelli, K.M., and Joshi, R. (2009). Hox specificity unique roles for co-factors and collaborators. *Curr. Top. Dev. Biol.* 88, 63–101.
- Meijnsing, S.H., Pufall, M.A., So, A.Y., Bates, D.L., Chen, L., and Yamamoto, K.R. (2009). DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* 324, 407–410.
- Meysman, P., Dang, T.H., Laukens, K., De Smet, R., Wu, Y., Marchal, K., and Engelen, K. (2011). Use of structural DNA properties for the prediction of transcription-factor binding sites in *Escherichia coli*. *Nucleic Acids Res.* 39, e6.
- Noro, B., Culi, J., McKay, D.J., Zhang, W., and Mann, R.S. (2006). Distinct functions of homeodomain-containing and homeodomain-less isoforms encoded by homothorax. *Genes Dev.* 20, 1636–1650.
- Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M., and Zhurkin, V.B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* 95, 11163–11168.
- Riley, T.R., Slattery, M., Abe, N., Rastogi, C., Liu, D., Mann, R.S., and Bussemaker, H.J. (2014). SELEX-seq: a method for characterizing the complete repertoire of binding site preferences for transcription factor complexes. *Methods Mol. Biol.* 1196, 255–278.
- Rohs, R., West, S.M., Liu, P., and Honig, B. (2009a). Nuance in the double-helix and its role in protein-DNA recognition. *Curr. Opin. Struct. Biol.* 19, 171–177.
- Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009b). The role of DNA shape in protein-DNA recognition. *Nature* 461, 1248–1253.
- Ryoo, H.D., and Mann, R.S. (1999). The control of trunk Hox specificity and activity by Extradenticle. *Genes Dev.* 13, 1704–1716.
- Seeman, N.C., Rosenberg, J.M., and Rich, A. (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA* 73, 804–808.
- Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., et al. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147, 1270–1282.
- Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordán, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* 39, 381–399.
- Weirauch, M.T., Cote, A., Norel, R., Annala, M., Zhao, Y., Riley, T.R., Saez-Rodriguez, J., Cokelaer, T., Vedenko, A., Talukder, S., et al.; DREAM5 Consortium (2013). Evaluation of methods for modeling transcription factor sequence specificity. *Nat. Biotechnol.* 31, 126–134.
- Yang, L., Zhou, T., Dror, I., Mathelier, A., Wasserman, W.W., Gordán, R., and Rohs, R. (2014). TFBSshape: a motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res.* 42, D148–D155.
- Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A.C., Ghane, T., Di Felice, R., and Rohs, R. (2013). DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale. *Nucleic Acids Res.* 41, W56–W62.
- Zhou, T., Shen, N., Yang, L., Abe, N., Horton, J., Mann, R.S., Bussemaker, H.J., Gordan, R., and Rohs, R. (2015). Quantitative modeling of transcription factor binding specificities using DNA shape. *Proc. Natl. Acad. Sci. USA*. Published online March 9, 2015. <http://dx.doi.org/10.1073/pnas.1422023112>.
- Zhu, L.J., Christensen, R.G., Kazemian, M., Hull, C.J., Enuameh, M.S., Basciotta, M.D., Brasefield, J.A., Zhu, C., Asriyan, Y., Lapointe, D.S., et al. (2011). FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* 39, D111–D117.

# The BRAF Pseudogene Functions as a Competitive Endogenous RNA and Induces Lymphoma In Vivo

## Graphical Abstract



## Authors

Florian A. Karreth, Markus Reschke, ..., Roberto Chiarle, Pier Paolo Pandolfi

## Correspondence

ppandolf@bidmc.harvard.edu

## In Brief

The in vivo evidence for the regulatory activity of pseudogenes has been lacking, and their role in disease progression has been correlative. This study now shows that transgenic expression of the BRAF pseudogene induces a malignancy in mice resembling human diffuse large B cell lymphoma, establishing its oncogenic function.

## Highlights

- The BRAF pseudogene functions as a ceRNA for BRAF in humans and mice
- *Braf-rs1* overexpression promotes B cell lymphoma in mice
- Silencing of *BRAFP1* affects MAPK signaling and proliferation of human cancer cells
- Genomic gains and aberrant expression of *BRAFP1* are found in various human cancers



# The BRAF Pseudogene Functions as a Competitive Endogenous RNA and Induces Lymphoma In Vivo

Florian A. Karreth,<sup>1</sup> Markus Reschke,<sup>1</sup> Anna Ruocco,<sup>1</sup> Christopher Ng,<sup>1</sup> Bjoern Chapuy,<sup>2</sup> Valentine Léopold,<sup>1</sup> Marcela Sjöberg,<sup>3</sup> Thomas M. Keane,<sup>3</sup> Akanksha Verma,<sup>4</sup> Ugo Ala,<sup>1</sup> Yvonne Tay,<sup>1,9</sup> David Wu,<sup>5</sup> Nina Seitzer,<sup>1</sup> Martin Del Castillo Velasco-Herrera,<sup>3</sup> Anne Bothmer,<sup>1</sup> Jacqueline Fung,<sup>1</sup> Fernanda Langellotto,<sup>6</sup> Scott J. Rodig,<sup>7</sup> Olivier Elemento,<sup>4</sup> Margaret A. Shipp,<sup>2</sup> David J. Adams,<sup>3</sup> Roberto Chiarle,<sup>6,8</sup> and Pier Paolo Pandolfi<sup>1,\*</sup>

<sup>1</sup>Cancer Research Institute, Beth Israel Deaconess Cancer Center, Department of Medicine and Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA

<sup>2</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

<sup>3</sup>Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1HH, UK

<sup>4</sup>Department of Physiology and Biophysics, Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY 10021, USA

<sup>5</sup>Meyer Cancer Center, Weill Cornell Medical College, New York, NY 10021, USA

<sup>6</sup>Department of Pathology, Children's Hospital and Harvard Medical School, Boston, MA 02115, USA

<sup>7</sup>Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA

<sup>8</sup>Department of Molecular Biotechnology and Health Sciences, University of Torino, 10124 Torino, Italy

<sup>9</sup>Present address: Cancer Science Institute of Singapore and Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597

\*Correspondence: [ppandolfi@bidmc.harvard.edu](mailto:ppandolfi@bidmc.harvard.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.043>

## SUMMARY

Research over the past decade has suggested important roles for pseudogenes in physiology and disease. In vitro experiments demonstrated that pseudogenes contribute to cell transformation through several mechanisms. However, in vivo evidence for a causal role of pseudogenes in cancer development is lacking. Here, we report that mice engineered to overexpress either the full-length murine *B-Raf* pseudogene *Braf-rs1* or its pseudo “CDS” or “3' UTR” develop an aggressive malignancy resembling human diffuse large B cell lymphoma. We show that *Braf-rs1* and its human ortholog, *BRAFP1*, elicit their oncogenic activity, at least in part, as competitive endogenous RNAs (ceRNAs) that elevate BRAF expression and MAPK activation in vitro and in vivo. Notably, we find that transcriptional or genomic aberrations of *BRAFP1* occur frequently in multiple human cancers, including B cell lymphomas. Our engineered mouse models demonstrate the oncogenic potential of pseudogenes and indicate that ceRNA-mediated microRNA sequestration may contribute to the development of cancer.

## INTRODUCTION

Over the past few years, remarkable progress has been made in establishing long non-coding RNAs (lncRNAs) as important regulators of various biological processes. Given their critical roles,

it is not surprising that aberrant expression and/or function of lncRNAs are implicated in the development of diseases such as cancer (Gutschner and Diederichs, 2012).

Pseudogenes, a sub-class of lncRNA genes that developed from protein-coding genes but have lost the ability to produce proteins, have long been viewed as non-functional genomic relicts of evolution (Poliseno, 2012). However, the vast majority of pseudogenes have protein-coding parental counterparts with which they share high sequence homology, which enables pseudogenes to participate in posttranscriptional regulation of their parental genes. Mechanisms of parental gene regulation include the formation of endogenous siRNAs (Tam et al., 2008; Watanabe et al., 2008), recruitment of regulatory proteins by pseudogene antisense RNAs to complementary sites in the parental gene to modulate chromatin remodeling and transcription (Hawkins and Morris, 2010; Johnsson et al., 2013), and competition for RNA-binding proteins or the translation machinery (Bier et al., 2009; Chiefari et al., 2010; Han et al., 2011).

We recently proposed that the high sequence homology enables pseudogenes to compete with their parental genes for a shared pool of common microRNAs (miRNAs) (Poliseno et al., 2010), thus regulating the latter's expression as competitive endogenous RNA (ceRNAs) (Salmena et al., 2011). This mechanism is of particular relevance to cancer where pseudogenes are aberrantly expressed (Kalyana-Sundaram et al., 2012). Specifically, we demonstrated that pseudogenes of the frequently mutated cancer genes *PTEN* and *KRAS* function as ceRNAs in vitro (Poliseno et al., 2010). Moreover, we and others reported that mRNAs and non-coding RNAs may serve as ceRNAs that regulate each other through miRNA-dependent crosstalk (Cazalla et al., 2010; Cesana et al., 2011; Franco-Zorrilla et al.,



2007; Hansen et al., 2013; Karreth et al., 2011; Libri et al., 2012; Marciniowski et al., 2012; Memczak et al., 2013; Sumazin et al., 2011; Tay et al., 2011; Wang et al., 2013), suggesting that pseudogenes regulate the expression of their parental genes in the context of larger networks of protein-coding and non-coding ceRNAs.

While sufficient data exist to demonstrate pseudogene functions in vitro, in vivo evidence for the regulatory activity of pseudogenes—either as ceRNAs or by any of the other above-mentioned mechanisms—is lacking, and their role in disease progression is correlative. Here, we describe a causal role for the *BRAF* pseudogene in the development of cancer.

## RESULTS

### The *BRAF* Pseudogene Regulates *BRAF* in a Dicer1-Dependent Manner

The *BRAF* pseudogene (*BRAFP1*) is overexpressed in various tumor types (Zou et al., 2009; Kalyana-Sundaram et al., 2012), suggesting that it may contribute to cancer development. We have shown that pseudogenes are able to regulate expression of their parental genes through sequestration of shared miRNAs (Poliseno et al., 2010), and *BRAFP1*-mediated elevation of *BRAF* may promote MAPK signaling and tumorigenesis. MiRNA predictions revealed that murine *Braf-rs1* (Gm18189) and *B-Raf* are targeted by 54 and 114 miRNA families, respectively, 53 of which they have in common. Similarly, human *BRAFP1* and *BRAF* are targeted by 60 and 48 miRNA families, respectively, and share 40 (Figures S1A–S1D, Table S1). Thus, the *BRAF* pseudogene may operate as a ceRNA for *BRAF* in mice and humans. Indeed, ectopic expression of *Braf-rs1* in NIH 3T3 fibroblasts and *BRAFP1* in human PC9 and HeLa cancer cells elevated *BRAF* protein and ERK phosphorylation (Figures 1A and S1E). Importantly, B-Raf was critical for this effect, as the *Braf-rs1*-induced increase in pERK was negated by genetic deletion of *B-Raf* in *B-Raf<sup>fl/fl</sup>* fibroblasts (Figure 1B). Moreover, expression of the *BRAF* pseudogene increased proliferation of NIH 3T3, PC9, and HeLa cells (Figures 1C, 1D, and S1F). Moderate B-Raf overexpression was sufficient to increase pERK expression, proliferation, and anchorage-independent growth of NIH 3T3 fibroblasts (Figures S1G–S1I), indicating that *Braf-rs1*-mediated elevation of B-Raf may be sufficient for the observed phenotype.

To test whether the effect of the *BRAF* pseudogene on *BRAF* expression and proliferation rates was dependent on miRNAs, we utilized cell lines lacking functional Dicer1, a ribonuclease critical for miRNA biogenesis and whose deficiency results in drastically reduced levels of mature miRNAs (Cummins et al., 2006; Ravi et al., 2012). Ectopic expression of *Braf-rs1* increased expression of B-Raf and pERK and elevated proliferation of Dicer1-proficient murine sarcoma cells, but not that of isogenic Dicer1 knockout cells (Figures 1E and 1F). Similarly, overexpression of *BRAFP1* in Dicer1-proficient human HCT116 colon cancer cells increased expression of *BRAF* and pERK and elevated proliferation, and these effects were abrogated in isogenic Dicer1 mutant HCT116 cells (Figures 1G and 1H). Thus, the *BRAF* pseudogene-induced effects are dependent on *BRAF* and Dicer1.

### The *BRAF* Pseudogene Regulates *BRAF* as a Competitive Endogenous RNA

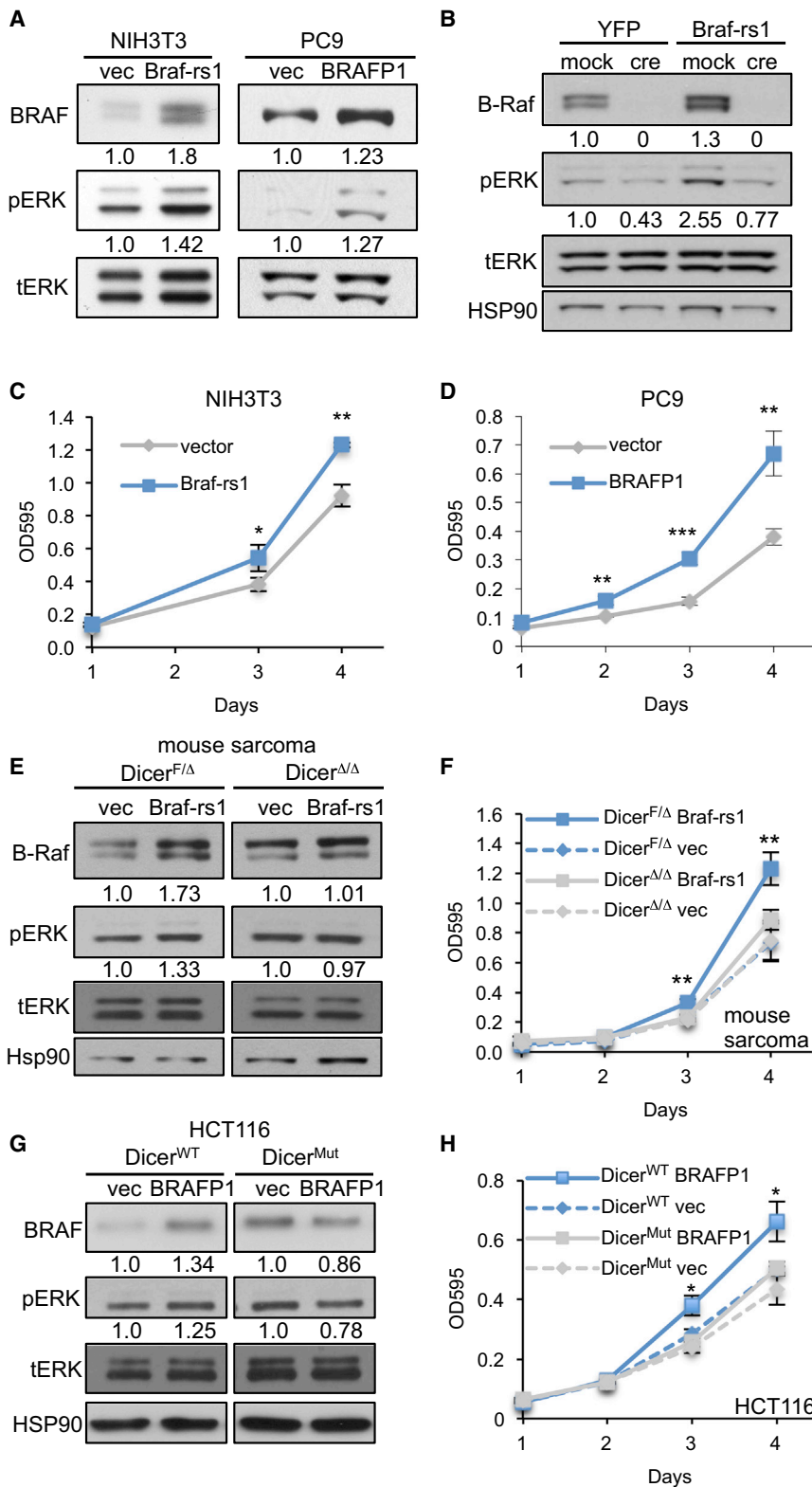
The finding that the *BRAF* pseudogene mediates its effect through mature miRNAs suggests that it may function as a ceRNA. To test this directly, we co-expressed *BRAFP1* with a human *BRAF*-3' UTR-luciferase reporter in Dicer1-proficient and -deficient HCT116 cells. *BRAFP1* elevated the activity of the *BRAF* 3' UTR-luciferase reporter in a Dicer1-dependent manner (Figure 2A), further supporting the notion that the cross-talk is mediated by mature miRNAs. To validate this result, we tested several predicted shared miRNAs in 3' UTR-luciferase reporter assays. Three out of ten murine miRNAs (miR-134, miR-543, and miR-653) significantly repressed *Braf-rs1* and *B-Raf* luciferase reporters (Figure 2B), suggesting that the cross-talk may be mediated at least in part by these three miRNAs.

Next, we determined the ability of *Braf-rs1* to decoy the dual targeting miRNAs miR-134, miR-543, and miR-653 from luciferase reporters carrying miRNA response elements (MREs). *Braf-rs1* regulated the expression of the luciferase reporters, especially at lower miRNA concentrations (Figure 2C). *Braf-rs1*-mediated sequestration of the least potent of the three dual targeting miRNAs, miR-543, had the most robust effect on luciferase reporter activity (Figure 2C). These data suggest that both potency and abundance of the miRNAs may be important determinants for ceRNA crosstalk. In addition, *Braf-rs1* was able to sequester endogenous miR-653, miR-134, and miR-543 from the respective luciferase-MRE reporters, and mutation of the MREs in *Braf-rs1* abrogated this effect (Figure 2D). Similarly, four out of nine human miRNAs (miR-30a, miR-182, miR-876, and miR-590) were able to repress *BRAF*- and *BRAFP1*-luciferase reporters (Figure S2A). miR-30a, miR-182, and miR-876 were also efficiently sequestered from the respective MRE-luciferase reporters by *BRAFP1*, and mutation of these miRNA-binding sites reduced *BRAFP1*'s activity as a miRNA sponge (Figure S2B).

### Generation of TRE-BPS Mice

As *Braf-rs1* regulates the expression of *B-Raf* and MAPK signaling, we sought to investigate whether aberrant *Braf-rs1* expression is oncogenic in vivo. To this end, we generated a transgenic allele containing murine *Braf-rs1* under the control of a doxycycline (Dox)-inducible Tet-response element (TRE) and targeted it to the collagen A1 locus using Flp recombinase-mediated genomic integration (Beard et al., 2006) (Figures S2C and S2D). We isolated mouse embryonic fibroblasts (MEFs) from TRE-*Braf-rs1* (henceforth referred to as TRE-BPS) mice to confirm that expression of the *Braf-rs1* allele regulates B-Raf. Infection of MEFs with a tTA-expressing retrovirus resulted in 6- to 18-fold induction of *Braf-rs1* expression (Figures 2E and S2E), as well as increased levels of B-Raf and pERK (Figure 2F) and proliferation (Figure 2G), confirming that the transgenic allele elicits effects similar to ectopic expression of *Braf-rs1*.

We used TRE-BPS MEFs to analyze the stoichiometry of *B-Raf* and *Braf-rs1*. First, we determined the absolute number of transcripts by qPCR using plasmids carrying *Braf-rs1* and *B-Raf* as standards (Figure S2E). In TRE-BPS MEFs infected with a control retrovirus, *B-Raf* molecules were 13- to 26-fold more abundant than *Braf-rs1*, while in tTA-infected cells, the *B-Raf*:*Braf-rs1* ratio



**Figure 1. The *BRAF* Pseudogene Regulates *BRAF* in a *Dicer1*-Dependent Manner**

(A) Western blot demonstrating increased BRAF and pERK expression upon ectopic *BRAF* pseudogene expression in mouse (NIH 3T3, left) and human (PC9, right) cells.

(B) Western blot of *B-Raf*<sup>fl/fl</sup> fibroblasts overexpressing *Braf-rs1* or control (yellow fluorescent protein [YFP]) in the presence or absence of Adeno-Cre infection.

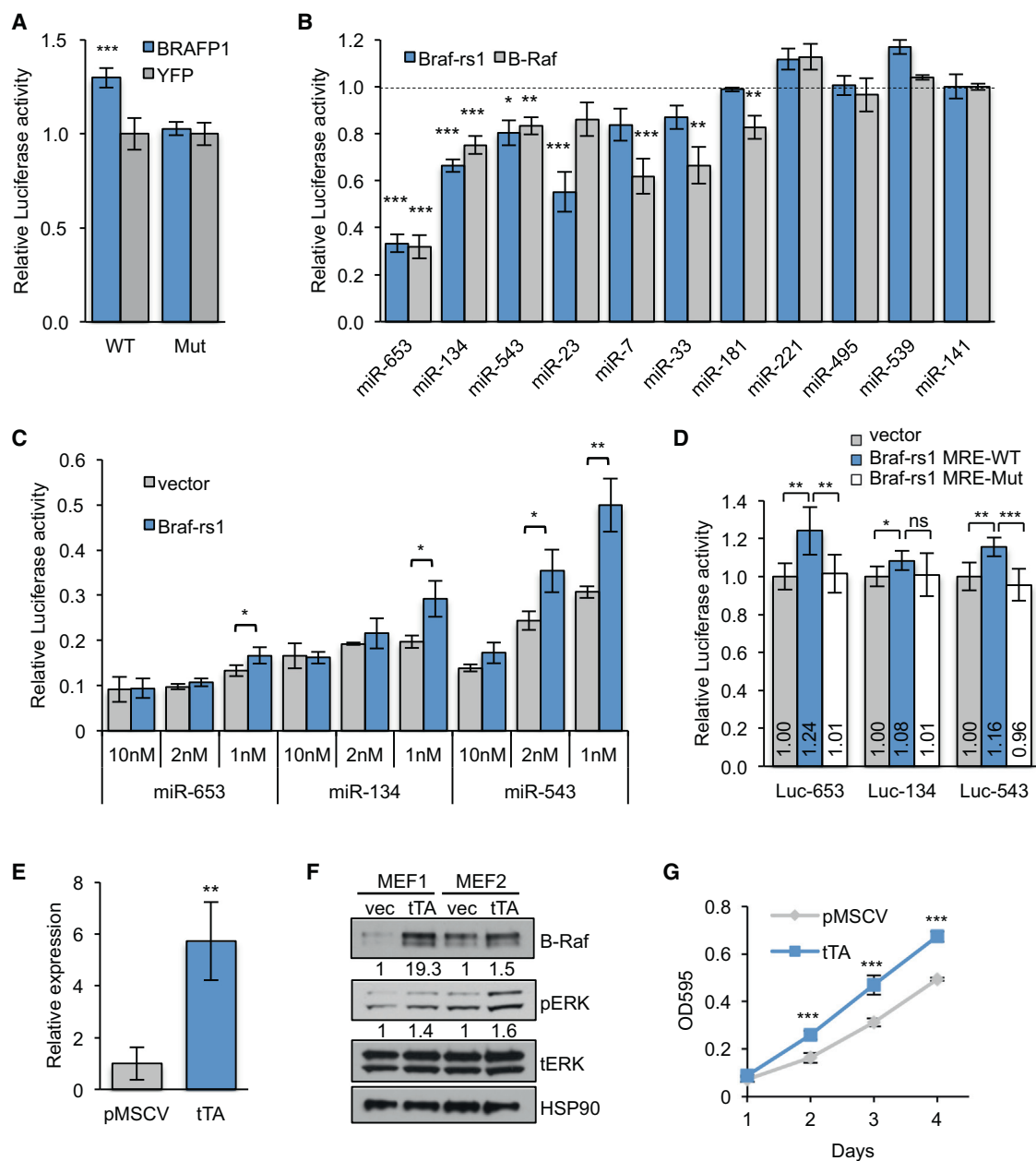
(C) Increased proliferation of NIH 3T3 fibroblasts upon ectopic *Braf-rs1* expression.

(D) Increased proliferation of PC9 cells upon ectopic *BRAF P1* expression.

(E and F) Western blot (E) and proliferation assay (F) of *Dicer1*<sup>FL/Δ</sup> and *Dicer1*<sup>Δ/Δ</sup> murine sarcoma cells overexpressing *Braf-rs1*.

(G and H) Western blot (G) and proliferation assay (H) of *Dicer1*<sup>WT</sup> and *Dicer1*<sup>Mut</sup> human HCT116 colon cancer cells overexpressing *BRAFP1*.

Error bars represent mean  $\pm$  SD. \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ . See also Figure S1.



**Figure 2. The *BRAF* Pseudogene Functions as a miRNA Sponge**

(A) *BRAF* 3' UTR-luciferase reporter assay in Dicer1<sup>WT</sup> and Dicer1<sup>Mut</sup> HCT116 cells expressing *BRAFP1* or control (YFP).

(B) Luciferase reporter assay using the 3' UTRs of *B-Raf* and *Braf-rs1* to analyze repression by the indicated miRNA mimics. miR141 serves as a negative control.

(C) *Braf-rs1* sequesters miRNAs to regulate MRE-Luc reporter activity. HEK293T cells were co-transfected with MRE-Luc reporter constructs, the respective miRNA mimics, and *Braf-rs1*-L277 or empty control L277 plasmids. The luciferase activity relative to a Luc reporter without MRE is shown.

(D) Luciferase activity measured in HEK293T cells co-expressing MRE-Luc reporters (Luc-653, Luc-134, or Luc-543) and wild-type or MRE mutant *Braf-rs1* or empty vector.

(E) qPCR showing tTA-induced *Braf-rs1* expression in TRE-BPS MEFs.

(F) Western blot for B-Raf and pERK in tTA-infected TRE-BPS MEFs.

(G) Proliferation of TRE-BPS MEF1 shown in (F).

Error bars represent mean  $\pm$  SD. \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ . See also Figure S2.

was between 1.3 and 2.5 (Figure S2E). RNA-sequencing (RNA-seq) analysis confirmed *Braf-rs1* induction and found *B-Raf*: *Braf-rs1* ratios in a range similar to that determined by qPCR (Fig-

ure S2F and data not shown). Next, we determined the number of molecules of miR-653, miR-134, and miR-543 in TRE-BPS MEFs by qPCR using standard curves. MiRNA expression was not

significantly affected upon transgene induction (Figure S2G, Table S2). miR-653 was expressed at extremely low levels, likely precluding it from *Braf-rs1/B-Raf* ceRNA crosstalk in MEFs. Additional predicted miRNAs that are expressed in MEFs (Table S2) but were not further validated may also contribute to crosstalk. Hence, the stoichiometry of *B-Raf*, transgenic *Braf-rs1*, and some dual-targeting miRNAs fits well within the optimal crosstalk criteria that we have recently established (Ala et al., 2013), supporting the hypothesis that overexpression of *Braf-rs1* increases B-Raf through its ceRNA activity.

### ***Braf-rs1* Causes Diffuse Large B Cell Lymphoma**

To induce global overexpression of *Braf-rs1* in vivo, TRE-BPS mice were crossed to CAG-rtTA3 mice (Prensrirut et al., 2011), and compound mutant animals and single mutant controls were placed on a Dox-containing diet at 3 weeks of age (Figure S3A). qPCR analysis after 4 weeks of Dox administration confirmed *Braf-rs1* overexpression in all organs tested (Figure S3B). Following 4 months of Dox treatment, TRE-BPS; CAG-rtTA3 mice became moribund and had to be sacrificed after a median survival of 421 days (Figure 3A), while none of the single-mutant animals or compound mutants maintained on a regular diet developed similar symptoms. All moribund TRE-BPS; CAG-rtTA3 mice presented with splenomegaly (Figures 3B and 3C) and enlarged lymph nodes (Figure 3K).

Histological analysis revealed large tumor nodules involving the splenic white pulp (Figures 3D and 3E). Tumors consisted of large lymphoid cells admixed with numerous plasmablasts and plasma cells (Figure 3F). The mitotic rate was very high (Figure 3F), and the proliferation rate was markedly increased compared to normal white pulp (Figures 3G and S3C).

We determined the immunophenotype of the splenic tumors by flow cytometry when TRE-BPS; CAG-rtTA3 mice succumbed to the malignancy. The cell population expressing surface B220 was decreased in spleens (Figure 3H), while Gr-1<sup>+</sup>/Mac-1<sup>+</sup> cells were slightly increased and CD3<sup>+</sup> cells were unchanged (Figures 3I and 3J). Lymph nodes displayed more B220<sup>+</sup> cells, while CD3<sup>+</sup> cells were less abundant (Figures 3L and 3M). Similar results were obtained when calculated as fold change relative to controls (Figures S3D–S3H). By immunohistochemistry, tumor cells stained positively for CD45R/B220 and IgG (Figures 4A and 4C) and negatively for CD3 (Figure 4B). Moreover, tumors were negative for the germinal center marker Bcl6 (Figure 4D) and strongly positive for Mum1 (Figure 4E), while residual germinal centers adjacent to the tumors were Bcl6 positive and Mum1 negative (Figures 4D and 4E). The decrease of B220 expression on the surface of tumor cells reflected the marked plasmacellular differentiation, as shown by the abundance of IgG<sup>+</sup> cells. Overall, this phenotype was consistent with post-germinal center diffuse large B cell lymphoma.

We next determined the abundance of *Braf-rs1*, *B-Raf*, and miRNA molecules in spleens after short-term Dox exposure (10 days) and in lymphomas and control spleens after long-term Dox exposure. While endogenous *Braf-rs1* expression was between 6- and 115-fold lower than *B-Raf*, expression of transgenic *Braf-rs1* was comparable to *B-Raf* (Figures S3I–S3L). Expression of miR-134, miR-543, and miR-653 was not affected by *Braf-rs1* overexpression (Figures S3M and S3N).

Similar to MEFs, miR-653 was expressed at low levels, while miR-134 and miR-543 were expressed at levels that are amenable to ceRNA crosstalk (Figures S3M and S3N).

### **Aggressive Lymphomas Are Transplantable and Depend on *Braf-rs1* Expression**

Macroscopic lymphoma nodules were commonly observed in the kidneys, livers, and lungs of TRE-BPS; CAG-rtTA3 mice (Figure 4F and data not shown), and histological analysis revealed microscopic organ infiltration by lymphoma cells in all animals (Figures 4G–4I). Such tumor cells displayed a CD45R/B220<sup>+</sup> and Mum1<sup>+</sup> phenotype identical to the cells infiltrating spleens and lymph nodes (Figures 4J–4O). Additionally, heterozygous loss of *Pten* reduced the median survival of TRE-BPS; CAG-rtTA3 mice to 172 days (data not shown).

To further assess the tumorigenicity of *Braf-rs1*-induced lymphomas, we analyzed their transplantation potential. NSG mice injected with TRE-BPS; CAG-rtTA3 spleen cells had to be sacrificed 100–150 days after transplantation due to deteriorating health. Moreover, NSG mice transplanted with TRE-BPS; CAG-rtTA3; *Pten*<sup>+/-</sup> lymphoma cells had to be sacrificed after 80 days (data not shown). NSG recipients exhibited infiltrating lymphoma cells in spleens, livers, lungs, and kidneys (Figure 5A). These results suggest that *Braf-rs1*-induced lymphomas are transplantable and highly aggressive.

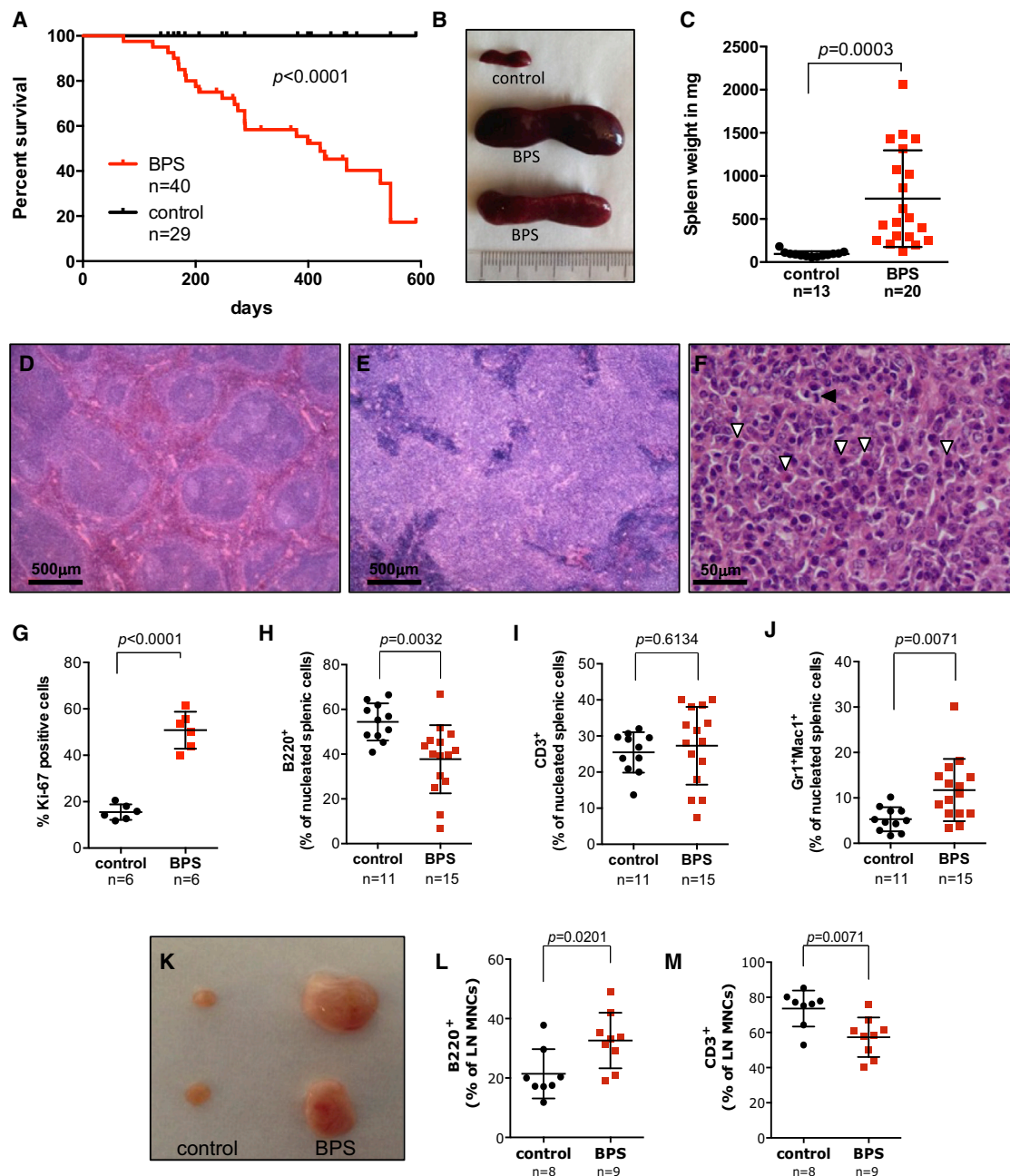
We next determined whether continuous expression of *Braf-rs1* was required for tumor maintenance. TRE-BPS; CAG-rtTA3 receiving a Dox-diet were monitored by palpation and were taken off Dox chow once splenomegaly became apparent. Spleen sizes of these animals were subsequently measured using high-resolution ultrasound. Notably, enlarged spleens of all TRE-BPS; CAG-rtTA3 mice reduced in size, while spleens of control mice were unaffected (Figure 5B). Moreover, 40 days after weaning the mice off Dox chow, the histology (Figures 5C and 5D) and Mum1 expression pattern (Figures 5E and 5F) of the white pulp of TRE-BPS; CAG-rtTA3 spleens were comparable to controls, confirming that lymphomas had largely regressed.

### ***Braf-rs1* Regulates B-Raf In Vivo**

To determine whether *Braf-rs1* functions as a ceRNA for *B-Raf* in vivo, we examined *Braf-rs1*-induced lymphomas for expression of B-Raf and pERK. Notably, *Braf-rs1*-induced lymphomas displayed increased levels of B-Raf and pERK (Figures 5G, 5H, and S4A) compared to adjacent normal white pulp. The difference in B-Raf and pERK levels between tumors and normal white pulp in the same mouse is likely due to positive selection of B cells that express the highest levels of *Braf-rs1*, B-Raf, and pERK.

We next analyzed whether MAPK signaling is critical for the growth of *Braf-rs1*-induced lymphomas. To this end, we treated NSG mice that were transplanted with *Braf-rs1*-induced lymphoma cells with the MEK inhibitor GSK1120212. Notably, treatment with GSK1120212 markedly impaired the ability of transplanted lymphomas to colonize the livers of NSG mice (Figure 5I). Moreover, Dox withdrawal reduced B-Raf and pERK expression in tumors, indicating that increased MAPK activation is stimulated by continuous *Braf-rs1* expression (Figure S4B). These data suggest that *Braf-rs1* elicits its oncogenic effects, at least in part, through B-Raf and the MAPK pathway.





**Figure 3. *Braf-rs1* Expression In Vivo Results in a Lymphoid Malignancy**

BPS, TRE-BPS; CAG-rTA3 mice on Dox; control, TRE-BPS, or CAG-rTA3 mice on Dox here and in all figures.

(A) Survival of BPS and control mice.

(B and C) Size (B) and weight (C) of BPS and control mouse spleens.

(D and E) Photomicrograph of a spleen from a control (D) and BPS mouse (E).

(F) Higher-magnification photomicrograph showing tumor cells in a BPS spleen. White arrowheads denote plasma cells, and black arrowhead highlights a mitotic figure.

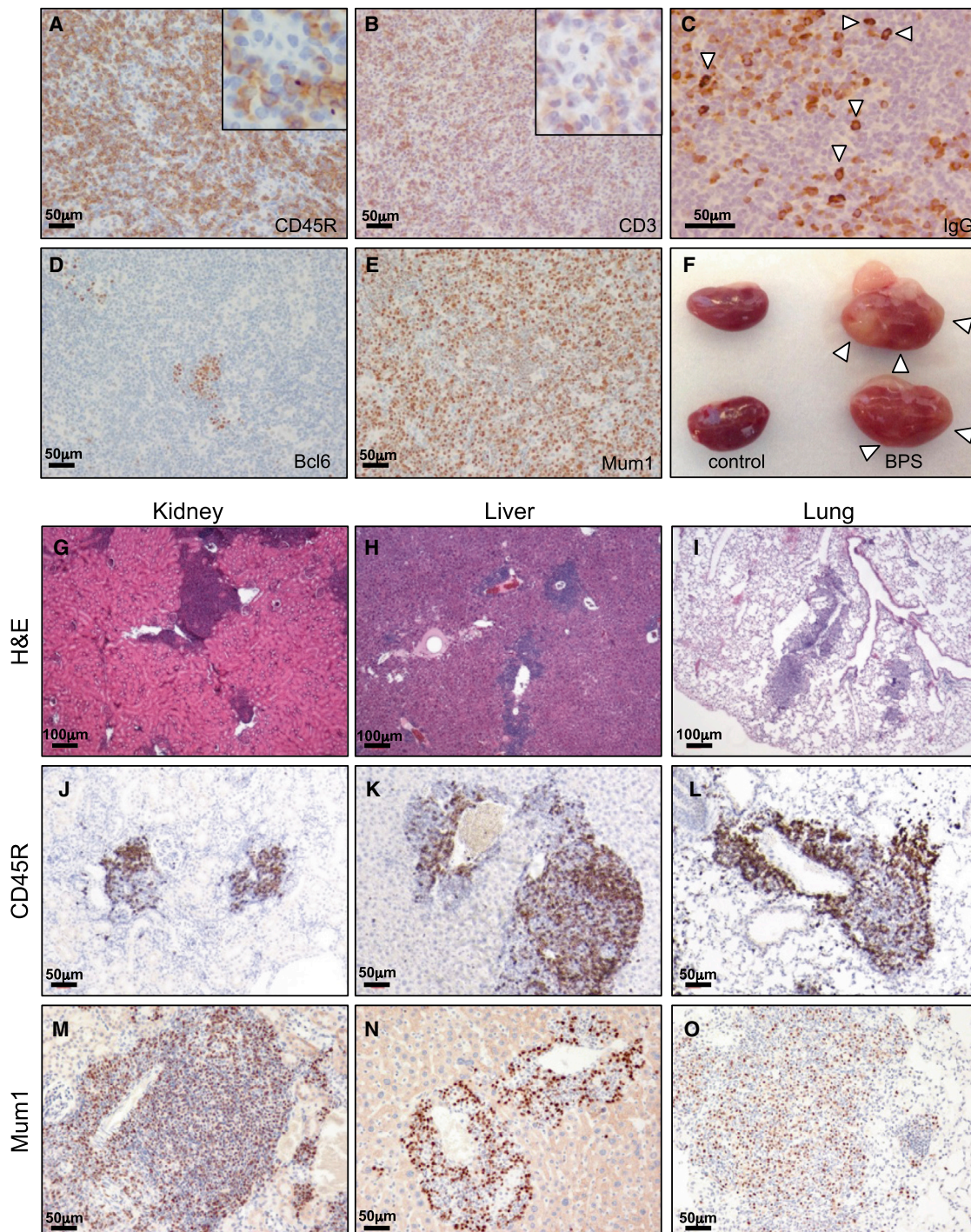
(G) Quantification of Ki-67 staining.

(H–J) Flow cytometry-based quantification of splenic B220<sup>+</sup> (H), CD3<sup>+</sup> (I), and Gr1<sup>+</sup>/Mac-1<sup>+</sup> (J) populations.

(K) Size of control and BPS mouse lymph nodes.

(L and M) Flow cytometry-based quantification of B220<sup>+</sup> (L) and CD3<sup>+</sup> (M) populations in lymph nodes.

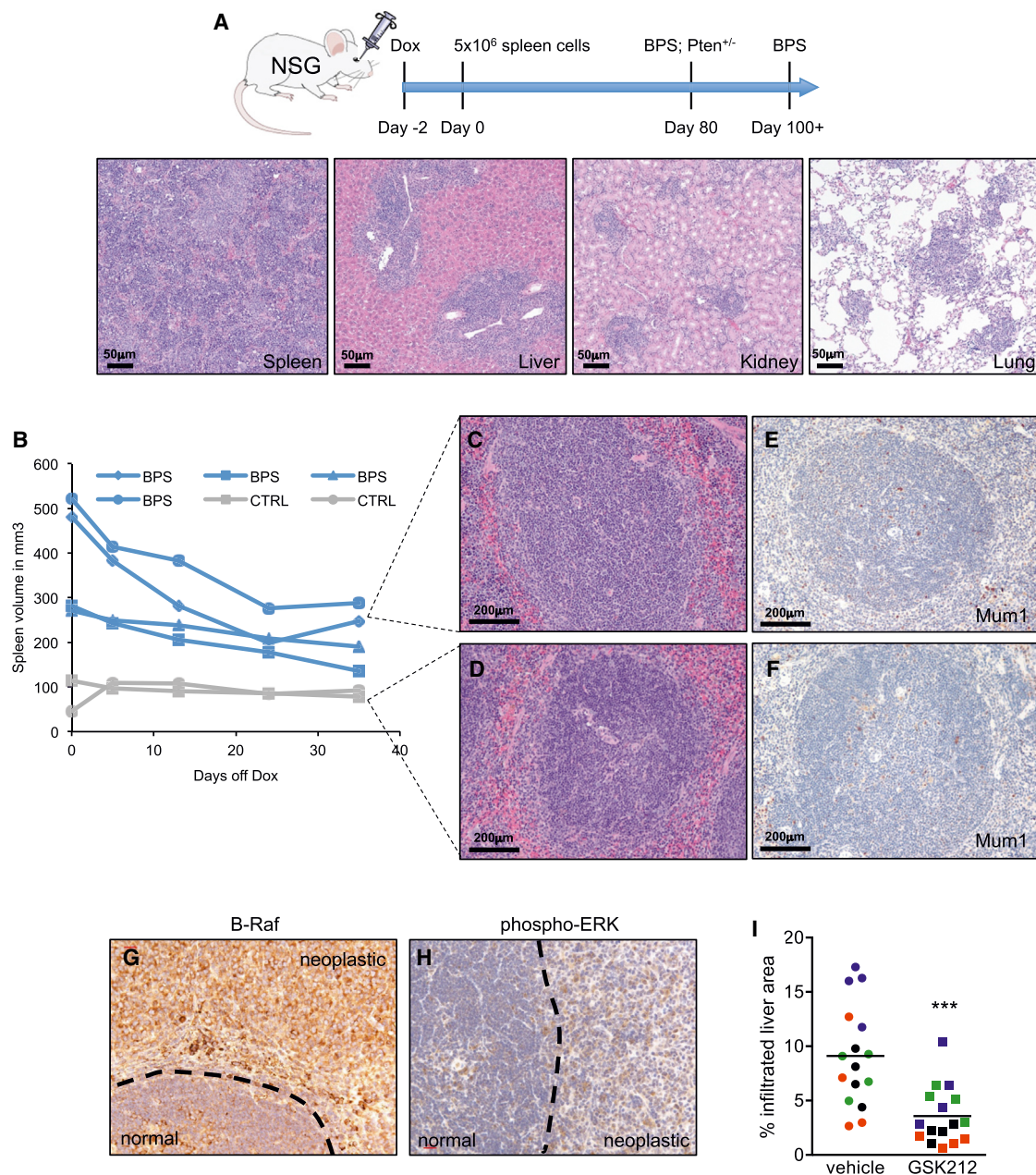
Error bars represent mean  $\pm$  SD. See also Figure S3.



**Figure 4. *Braf-rs1* Induces Diffuse Large B Cell Lymphoma**

(A) CD45R/B220 staining. Higher magnification inset shows staining of large lymphoma cells.  
 (B) CD3 staining. Higher-magnification inset shows positive staining of reactive T cells.  
 (C) IgG staining. Arrowheads denote plasma cells.  
 (D) Bcl6 staining. Lymphoma cells are negative, and residual germinal center is positive.  
 (E) Mum1 staining. Tumor cells are positive, and residual germinal center is negative.  
 (F) Photograph of control and BPS kidneys. Arrowheads denote tumor nodules.  
 (G–I) H&E staining of kidney (G), liver (H), and lung (I) sections from BPS mice.  
 (J–L) CD45R/B220 immunohistochemistry of kidney (J), liver (K), and lung (L) sections from BPS mice.  
 (M–O) Mum1 immunohistochemistry of kidney (M), liver (N), and lung (O) sections from BPS mice.





**Figure 5. Lymphomas Are Transplantable, Are Addicted to *Braf-rs1* Expression, and Activate the MAPK Pathway**

(A) Transplanted lymphoma cells infiltrating the spleen, liver, kidney, and lungs of NSG recipient mice.

(B) Spleen size measurements after Dox withdrawal.

(C–F) H&E staining (C and D) and Mum1 immunohistochemistry (E and F) of BPS and control mouse spleens depicted in (B) after Dox withdrawal.

(G) Immunohistochemical staining for B-Raf of lymphoma and adjacent normal white pulp in BPS spleen.

(H) Immunohistochemical staining for pERK of lymphoma and adjacent normal white pulp in BPS spleen.

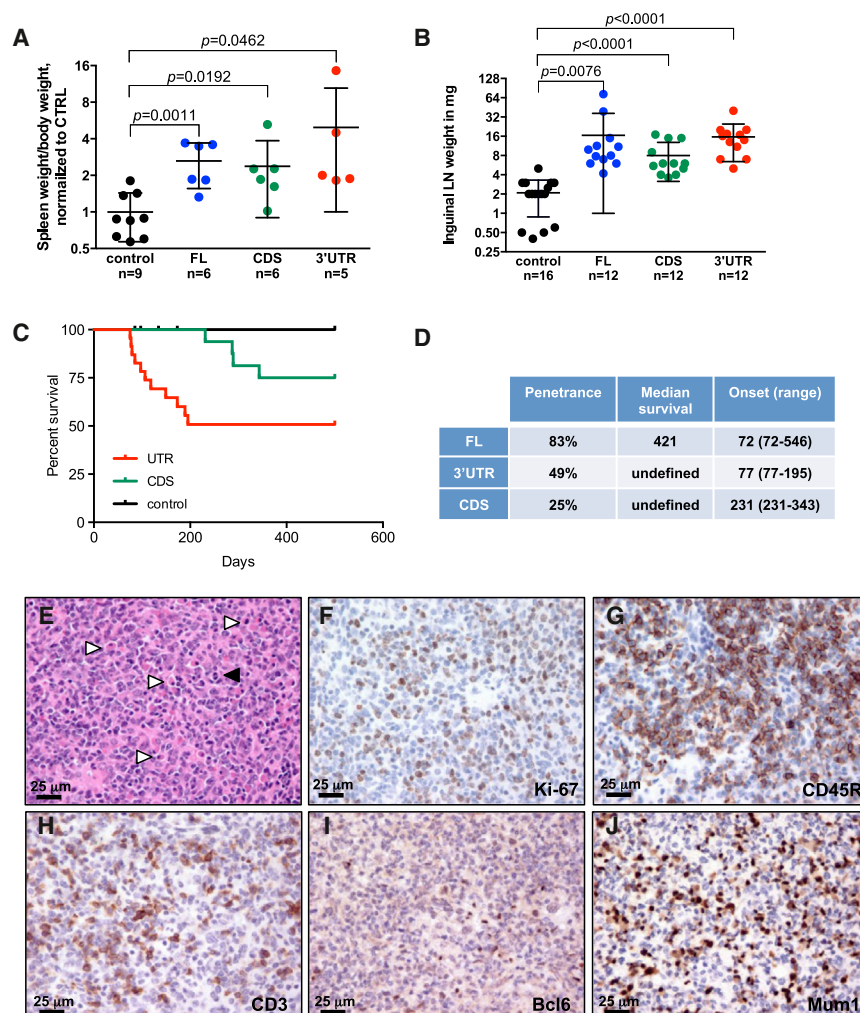
(I) Percentage of liver infiltration by TRE-BPS; CAG-rTA3; Pten<sup>+/-</sup> lymphoma cells transplanted into NSG mice in response to GSK1120212 treatment. Each symbol represents a liver section, and each recipient mouse is color coded.

Error bars represent mean ± SD. \*\*\*p ≤ 0.001. See also Figure S4.

### The “CDS” and “3′ UTR” of *Braf-rs1* Possess Oncogenic Potential

Based on *Braf-rs1*’s ability to decoy miRNAs, we reasoned that shorter fragments of *Braf-rs1* may be able to crosstalk with

*B-Raf* through a subset of the shared miRNA pool. Such fragments would elicit similar phenotypes provided that the crosstalk remains robust. Alternatively, different portions of *Braf-rs1* may regulate distinct ceRNA networks and yield distinct,



**Figure 6. *Braf-rs1*<sup>CDS</sup> and *Braf-rs1*<sup>3'UTR</sup> Possess Oncogenic ceRNA Activity Similar to Full-Length *Braf-rs1***

(A and B) Weights of spleens (A) and inguinal lymph nodes (B) of the indicated mouse strains after 6 months on Dox.

(C) Survival of TRE-BPS<sup>3'UTR</sup> and TRE-BPS<sup>CDS</sup> mice.

(D) Table summarizing the penetrance, median survival, and disease onset of TRE-BPS, TRE-BPS<sup>3'UTR</sup>, and TRE-BPS<sup>CDS</sup> mice.

(E) H&E staining of *Braf-rs1*<sup>3'UTR</sup>-induced lymphoma. White arrowheads indicate plasma cells, and black arrowhead indicates mitotic figure.

(F–J) Immunohistochemical staining of *Braf-rs1*<sup>3'UTR</sup>-induced lymphoma for Ki-67 (F), CD45R/B220 (G), CD3 (H), Bcl6 (I), and Mum1 (J).

Error bars represent mean  $\pm$  SD. See also Figure S5.

and proliferation, while the *Braf-rs1*<sup>CDS</sup>-induced effects were negligible (Figures S5D–S5G). *Braf-rs1*<sup>CDS</sup> and *Braf-rs1*<sup>3'UTR</sup> may regulate distinct ceRNA networks, but the finding that the severity of the phenotype elicited by the three *Braf-rs1* variants correlated with their ability to deregulate B-Raf provides compelling support to the notion that *Braf-rs1* operates as a proto-oncogenic ceRNA through B-Raf in B cells.

### **BRAFP1 Is an Oncogenic ceRNA in Human Cancer**

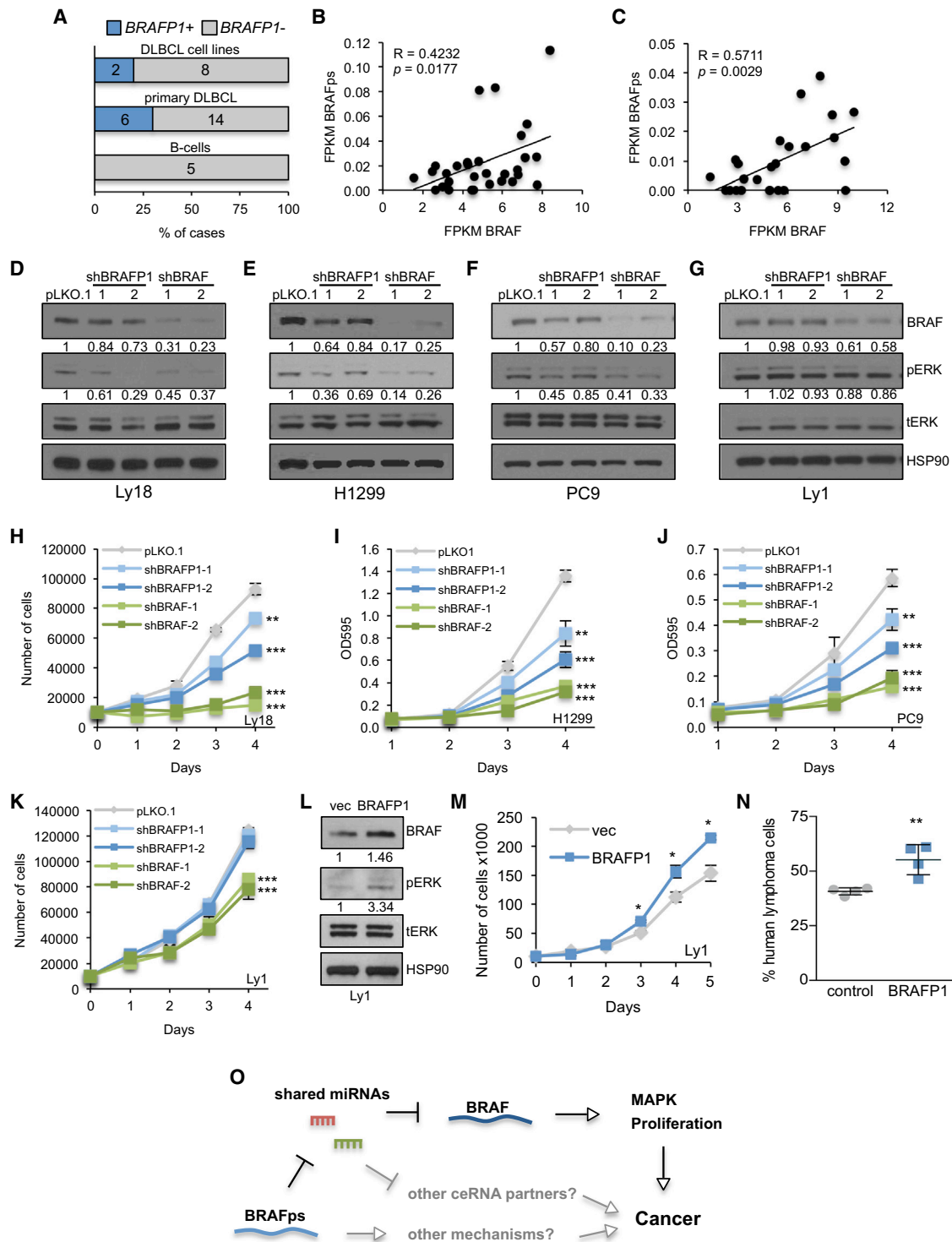
Overexpression of human *BRAFP1* increased BRAF and pERK levels as well as proliferation of human cells (Figures

1A, 1D, 1G, and 1H), suggesting that *BRAFP1* may be an oncogene in human cancer. To explore this possibility further, we first determined whether *BRAFP1* is expressed in human DLBCL. Interestingly, *BRAFP1* expression was not found in primary human B cells (Figures 7A and S6A) but was detected in 30% of human primary DLBCL and 20% of human DLBCL cell lines (Figures 7A and S6A). Similar observations have been made in the thyroid, where *BRAFP1* was expressed in some tumors, but not in normal tissue (Zou et al., 2009). Moreover, *BRAFP1* was expressed in melanoma, prostate cancer, and lung cancer cell lines (Figure S6A).

We next interrogated The Cancer Genome Atlas's (TCGA) cBio Cancer Genomics Portal for genomic abnormalities of the locus containing *BRAFP1*. As pseudogene data are not yet included in TCGA, we focused our analysis on two protein-coding genes flanking *BRAFP1*: *ZDHC15* and *MAGEE2* (Figure S6B). Notably, concurrent copy-number gains and amplification of *ZDHC15* and *MAGEE2* were observed in numerous cancer types (Figure S6B). Importantly, *BRAFP1* expression could be detected in such cancer types (Kalyana-Sundaram et al., 2012). Thus, both transcriptional mechanisms and genomic

abnormalities may contribute to the oncogenic activity of *BRAFP1*. To experimentally examine these possibilities, we generated two additional Dox-inducible mouse models overexpressing either the “CDS” or the “3' UTR” of *Braf-rs1* (Figures S2C and S2D). TRE-BPS<sup>CDS</sup> and TRE-BPS<sup>3'UTR</sup> mice were crossed to CAG-rTA3 mice and their offspring fed a Dox-containing diet for 6 months. Remarkably, both TRE-BPS<sup>CDS</sup> and TRE-BPS<sup>3'UTR</sup> mice displayed enlarged spleens and lymph nodes similar to full-length TRE-BPS mice (Figures 6A and 6B). *Braf-rs1*<sup>3'UTR</sup> overexpression resulted in splenomegaly and reduced survival (Figures 6C and 6D and S5C) similar to TRE-BPS mice. The histology and immunophenotype of lymphomas in TRE-BPS<sup>3'UTR</sup> mice were similar to that of full-length TRE-BPS animals (Figures 6E–6J, S5A, and S5B), indicating that *Braf-rs1*<sup>3'UTR</sup> overexpression elicits a phenotype similar to full-length *Braf-rs1*. TRE-BPS<sup>CDS</sup> mice developed lymphomas with a reduced penetrance and aggressiveness compared to mice overexpressing full-length *Braf-rs1* or *Braf-rs1*<sup>3'UTR</sup> (Figures 6C and 6D and data not shown). Similarly, infection of TRE-BPS<sup>CDS</sup> and TRE-BPS<sup>3'UTR</sup> MEFs with tTA-pMSCV induced *Braf-rs1*<sup>CDS</sup> and *Braf-rs1*<sup>3'UTR</sup> expression (Figure S5D), but only *Braf-rs1*<sup>3'UTR</sup> elicited a significant effect on B-Raf expression





**Figure 7. BRAFP1 in Human Cancer**

(A) Percentage of primary human B cells, primary human DLBCL, and human DLBCL cell lines expressing *BRAFP1* as determined by qPCR analysis. (B and C) Positive correlation of *BRAFP1* and *BRAF* expression in human DLBCL primary tumors (B) and cell lines (C). (D–G) Western blot for *BRAF* and *pERK* in OCI-Ly18 (D), H1299 (E), PC9 (F), and OCI-Ly1 (G) cells in response to *BRAFP1* silencing. (H–K) Proliferation of OCI-Ly18 (H), H1299 (I), PC9 (J), and OCI-Ly1 (K) cells in response to *BRAFP1* silencing. (L) Western blot for *BRAF* and *pERK* in OCI-Ly1 cells overexpressing *BRAFP1*.

(legend continued on next page)

aberrations may lead to abnormal *BRAFP1* expression in human cancer.

Our experiments in human cell lines indicate that *BRAFP1* may operate as a ceRNA to regulate BRAF expression. Accordingly, analysis of RNA sequencing data revealed that *BRAFP1* and *BRAF* expression were positively correlated in primary human DLBCL tumors and DLBCL cell lines (Figure 7B and 7C). We also analyzed whether the expression of dual-targeting miRNAs correlates with *BRAF* and/or *BRAFP1* expression. While miR-590 expression negatively correlated with *BRAFP1* levels, miR-30a, miR-182, and miR-876 showed no correlation (Figure S6C). Thus, similar to our observations in TRE-BPS MEFs, expression of *BRAFP1* and *BRAF* may not affect miRNA abundance in human DLBCL.

To functionally validate the oncogenic function of *BRAFP1* in human cancer, we designed shRNAs to specifically silence expression of endogenous *BRAFP1* (Figure S6H). Knockdown of *BRAFP1* in OCI-Ly18 DLBCL cells and H1299 and PC9 lung cancer cells reduced the expression of BRAF and pERK (Figures 7D–7F and S6I–S6K). *BRAFP1* silencing moderately reduced *BRAF* mRNA levels in OCI-Ly18 and PC9 cells, but not in H1299 cells, suggesting that the mechanism of miRNA-mediated regulation of *BRAF* varies between cell lines. Importantly, the *BRAFP1* hairpins had no effect on BRAF and pERK expression in OCI-Ly1 DLBCL cells that do not express endogenous *BRAFP1* (Figure 7G). Moreover, *BRAFP1* silencing reduced proliferation of OCI-Ly18, H1299, and PC9 cells, but not of OCI-Ly1 cells (Figures 7H–7K). Remarkably, silencing of endogenous *BRAFP1* elicited a significant effect on BRAF expression in OCI-Ly18, H1299, and PC9 cells even though it is ~15- to ~30-fold less abundant than *BRAF* (Figures S6D and S6E). Intriguingly, *BRAFP1* was turned over significantly faster than *BRAF* (Figure S6F), suggesting that the relatively low expression levels of *BRAFP1* may be due to its short half-life. We also determined the abundance of miR-30a, miR-182, and miR-876 in OCI-Ly18, H1299, and PC9 cells and found that their expression levels were in the same range as those of *BRAFP1* and *BRAF* (Figures S6G).

Overexpression of *BRAFP1* in three human DLBCL cell lines lacking endogenous *BRAFP1* expression, SU-DHL-4, Karpas422, and OCI-Ly1 (Figures S6A and S6L), resulted in elevated BRAF and pERK levels (Figures 7L and S6M). Moreover, *BRAFP1* overexpression increased proliferation of all three DLBCL cell lines (Figures 7M, S6N, and S6O) and resulted in increased growth of xenotransplanted OCI-Ly1 cells in the bone marrow of NSG recipients (Figure 7N). These data suggest that *BRAFP1* has oncogenic properties in human cancer.

## DISCUSSION

We investigated whether pseudogenes exert critical functions in the context of a whole organism and whether their perturbation contributes to the development of disease. We focused on the

*BRAF* pseudogene, as it exists in humans and mice and is downregulated in cancer (Kalyana-Sundaram et al., 2012; Zou et al., 2009). Our study establishes the *BRAF* pseudogene as a potent proto-oncogene that can elicit a phenotype resembling human diffuse large B cell lymphoma. Remarkably, no additional engineered mutations were required to drive this phenotype, and lymphomas completely regressed upon Dox withdrawal, emphasizing the oncogenic potential of the *BRAF* pseudogene. While it is possible that the *BRAF* pseudogene elicits its effects through more than one mechanism or pathway, the fact that both the CDS and the 3' UTR of *Braf-rs1* displayed a similar phenotype to full-length *Braf-rs1*, albeit with different severity, supports the notion that *Braf-rs1* functions as a ceRNA to regulate B-Raf in vivo (Figure 7O). Whether the oncogenic activity of *Braf-rs1* also requires additional ceRNA targets or non-ceRNA-related mechanisms will be the focus of future studies.

Several groups, including ours, have generated mathematical models to quantitatively assess the response of a ceRNA network to perturbations (Ala et al., 2013; Bosia et al., 2013; Figliuzzi et al., 2013). More recently, such models were used in conjunction with miRNA predictions, RNA sequencing, and target site occupancy analyses to more accurately characterize miRNA competition (Bosson et al., 2014; Denzler et al., 2014; Jens and Rajewsky, 2014). Intriguingly, these studies yielded disparate conclusions. It was proposed that ceRNA crosstalk is unlikely to occur upon physiological changes of ceRNA expression based on these models' estimates of the number of additional target sites required to achieve significant expression changes of other targets (Denzler et al., 2014; Jens and Rajewsky, 2014). By contrast, using Argonaute iCLIP and RNA-seq, Sharp and colleagues determined that a relatively low number of additional target sites could elicit ceRNA crosstalk when the number of miRNA molecules and high-affinity target sites approaches equimolarity (Bosson et al., 2014). Interestingly, *BRAFP1* is several-fold less abundant than *BRAF*, yet its silencing significantly diminished BRAF expression levels, MAPK signaling, and proliferation. *BRAF* and its pseudogene harbor high-affinity sites for the murine and human miRNAs that we validated as potential mediators of the ceRNA crosstalk (miRs-134, -543, and -653 and miRs-30a, -182, -876, respectively). Notably, the levels of these miRNAs in mouse spleens and lymphomas as well as human cancer cell lines are amenable to miRNA competition in accordance with the model proposed by Bosson et al. Thus, a ceRNA effect of *BRAFP1* that is solely based on miRNA competition may be compatible with this model.

Importantly, the studies by the groups of Sharp, Stoffel, and Rajewsky focused on ceRNA regulation that is mediated by a single miRNA. However, ceRNA pairs in general, and gene/pseudogene pairs in particular, share numerous miRNAs. This increases the likelihood of shared miRNAs being present at cross-talk-favoring levels, and we have shown that ceRNA crosstalk is enhanced when it is mediated by more miRNAs (Ala et al., 2013).

(M) Proliferation of OCI-Ly1 cells.

(N) Percentage of human CD19<sup>+</sup> transplanted OCI-Ly1 cells in bone marrow of NSG recipients.

(O) Model depicting the proposed oncogenic action of the *BRAF* pseudogene.

Error bars represent mean  $\pm$  SD. \* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ . See also Figure S6.

As discussed by Jens and Rajewsky, several factors that may influence ceRNA crosstalk are neglected in current mathematical models. For instance, subcellular co-localization of miRNAs and competing targets may result in local concentrations that favor ceRNA crosstalk. In addition, target degradation may trap miRNAs in P bodies or other sites of RNA decay, thus amplifying the ceRNA regulation by removing miRNAs from the available pool. Intriguingly, *BRAFP1* is degraded significantly faster than *BRAF* (Figure S6F); however, whether this influences the ceRNA activity of *BRAFP1* remains to be determined. Future improvements to both quantitative measurements and mathematical models will undoubtedly provide a better understanding of the molecular conditions required for ceRNA crosstalk. However, it should be noted that ceRNA crosstalk can be predicted solely based on the MRE overlap of transcripts (Chiu et al., 2014; Karreth et al., 2011; Sumazin et al., 2011; Tay et al., 2011), suggesting that miRNA competition is indeed the central component of ceRNA crosstalk.

Human hematopoietic malignancies are associated with “overdosage” of the X chromosome, which harbors the *BRAF* pseudogene locus. This can occur through *XIST* deletion and X chromosome duplication in women with myeloid cancers, and extra X chromosomes have been noted in a variety of hematopoietic cancers of both sexes (Dewald et al., 1989; Dierlamm et al., 1995; Heinonen et al., 1999; Paulsson et al., 2010; Rack et al., 1994; Yamamoto et al., 2002), including DLBCL (Bea et al., 2005; Monni et al., 1996; Morin et al., 2013). Our analysis revealed that a variety of human cancers harbor copy-number gains and amplifications of the locus containing *BRAFP1*. It is therefore tempting to speculate that increased X dosage and the potentially associated overexpression of *BRAFP1* contribute to the development and/or progression of cancer cases harboring more than one active copy of the X chromosome. Moreover, elevated expression of *BRAFP1* has been observed in cancers other than DLBCL (Kalyana-Sundaram et al., 2012; Zou et al., 2009), and transcriptional deregulation may thus be another means to deregulate *BRAFP1* expression. Whether *BRAFP1* has oncogenic potential in other organs such as the thyroid remains to be determined through the use of tissue-specific overexpression of the *BRAF* pseudogene.

Interestingly, several observations suggest that the *BRAF* pseudogenes evolved independently in mice and humans. First, they reside in non-syngeneic locations—on chromosome 10 in mice and on the X chromosome in humans. Second, the 3′ UTR of the *BRAF* gene is not conserved between mice and humans; importantly, however, the *BRAF* pseudogene 3′ UTRs display high sequence homology to their parental counterparts in the respective species. Third, murine *Braf-rs1* arose from an alternative *B-Raf* splice form that is specific to mice (Karreth et al., 2009). The likely parallel yet converging evolution of *BRAFP1* and *Braf-rs1* and the fact that the gene-pseudogene crosstalk is mediated by different miRNAs in the two species suggest that their functions may be conserved. Indeed, the frequent *BRAFP1* copy-number gains and transcriptional activation of *BRAFP1* in human cancers as well as our silencing and overexpression experiments indicate that our findings in the mouse are of relevance to human disease.

It was recently proposed that human *BRAFP1* encodes a peptide with the ability to activate the MAPK pathway (Zou et al., 2009). We neither detected any peptide translation by the mouse or human *BRAF* pseudogenes nor could we detect robust association of *Braf-rs1* with actively translating ribosomes (data not shown). These findings suggest that *Braf-rs1* is not translated into an oncogenic peptide but, rather, exerts its function as a RNA transcript. This is further supported by the finding that TRE-BPS<sup>3′UTR</sup> mice display a more severe phenotype compared to TRE-BPS<sup>CDS</sup> mice, which suggests that the effects of *Braf-rs1* on *B-Raf* are primarily mediated through its 3′ UTR. The *BRAFP1* ORF predicted by Zou et al., however, localizes to the CDS portion of the pseudogene.

Pseudogenes were considered genomic junk for decades, but their retention during evolution argues that they may possess important functions and that their deregulation could contribute to the development of disease. Indeed, several lines of evidence have associated pseudogenes with cellular transformation (Poli-seno, 2012). Our study shows that aberrant expression of a pseudogene causes cancer, thus vastly expanding the number of genes that may be involved in this disease. Moreover, our work emphasizes the functional importance of the non-coding dimension of the transcriptome and should stimulate further studies of the role of pseudogenes in the development of disease.

## EXPERIMENTAL PROCEDURES

### Flow Cytometry

Mice were euthanized and single-cell suspensions from spleens and lymph nodes were prepared by passing organs through 100 μm cell strainers in 2% FBS/PBS, centrifuged and re-suspended in 1–2 ml ACK red cell lysis buffer (GIBCO). Red blood cells were lysed on ice for 1 min. Cell suspensions were then washed in 2% FBS/PBS, centrifuged and re-suspended in 1 ml 2% FBS/PBS. For hematopoietic lineage analysis, we used monoclonal antibodies specific for the following: CD3e-PE (145-2C11), B220-FITC (RA3-6B2), Gr-1-APC (RB6-8C5), and CD11b-PE/Cy7 (M1/70). All antibodies were from eBioscience. To assess cell viability, cells were incubated with DAPI prior to FACS analysis. All staining mixtures were analyzed on a BD LSR II flow cytometer (Becton Dickinson). Resulting profiles were further processed and analyzed using the FlowJo 8.7 software. For fold change quantifications, both mutant and control cell populations were normalized to the average of the controls. At least five mice from different litters were used for all flow cytometry experiments.

### Tissue Fixation, H&E, and IHC

Tissues were fixed in 4% paraformaldehyde overnight and embedded in paraffin according to standard procedures. 5 μm sections were either stained with hematoxylin & eosin or with the following antibodies: CD45R/B220 (ab64100, Abcam), CD3 (ab5690, Abcam), Ki-67 (RM-9106-S1, Thermo Scientific), IgG (BA2000, Vector), BRAF (sc-9002, Santa Cruz), pERK (4373, Cell Signaling), Bcl-6 (5650, Cell Signaling), and Mum1 (sc-6059, Santa Cruz). Organs from at least five mice from different litters were used for all stainings.

### Cell Culture

HCT116 and HeLa were from ATCC, Dicer mutant HCT116 cells were provided by B. Vogelstein, and *Dicer<sup>FL/Δ</sup>* and *Dicer<sup>Δ/Δ</sup>* mouse sarcoma cells were provided by P. Sharp and were cultured in DMEM containing 10% FCS and 2 mM L-glutamine. PC9, H1299, H441, and H2009 (all provided by L. Cantley), OCI-Ly8, OCI-Ly3, RCK8, and Val were grown in RPMI-1640 containing 10% FCS and 2 mM L-glutamine. SU-DHL-4, SU-DHL-8, Karpas422, OCI-Ly7, Toledo, OCI-Ly1, and OCI-Ly18 cells were grown as previously described

(Chapuy et al., 2013). Cells were regularly tested with MycoAlert (Lonza) to ascertain that cells were not infected with mycoplasma.

### Plasmids, Transfection, and Virus Infection

Human *BRAF*P1 was cloned into pLenti-CMV-GFP-Puro (Addgene 25873) and pCDNA3, and mouse *Braf-rs1* was cloned into pCCL.sin.PPT.hPGK.GFP.Wpre (L277, L. Naldini) or pCDNA3-neo. pMSCV-tTA (Addgene #18783) was used to induce *Braf-rs1* expression in TRE-BPS MEFs. Lipofectamine 2000 was used for plasmid transfection. Lentivirus or retrovirus was produced in HEK293T LentiX cells (Clontech) co-transfected with VSVG, pMDL, and Rev or Eco helper plasmids, respectively. Viral supernatants were filtered and cells infected in the presence of 5  $\mu$ g/ml polybrene.

### Proliferation Assays

For proliferation assays,  $2 \times 10^4$  cells were plated in four 12-well plates in triplicates. Every day, one plate was fixed with 4% paraformaldehyde and stained with Crystal Violet. The dye was extracted with 10% acetic acid and its absorbance determined at OD<sub>595</sub>. For suspension cells,  $1 \times 10^4$  cells were plated in triplicates in round-bottom 96-well plates and counted every day for 5 days.

### Luciferase Assays

HCT116 cells were transfected with 150 ng of psiCHECK2 vector or psiCHECK2-humanBRAF 3' UTR and 1 mg human *BRAF*P1 constructs using Lipofectamine 2000. To validate miRNA targeting, 3' UTRs of murine and human gene and pseudogene were cloned into psiCHECK2.  $5 \times 10^4$  HEK293T cells were transfected in 48-well plates with 20 ng of psiCHECK2 reporter and 100 nM miRNA mimic (QIAGEN). To test the ceRNA activity of the *BRAF* pseudogenes,  $5 \times 10^4$  HEK293T cells were transfected in 48-well plates with 20 ng of psiCHECK2 reporter and 250 ng of murine *Braf-rs1*-L277 vector or human *BRFAP1*-pCDNA3 and 1–2 nM miRNA mimic. In all transfections, firefly luciferase activity was used as a normalization control for transfection efficiency. 48 hr after transfection, luciferase activities were measured consecutively with the dual luciferase reporter system (Promega).

### Western Blot

Cells were lysed in RIPA buffer containing HALT protease and phosphatase inhibitors (Sigma). 20  $\mu$ g total protein were separated on 4%–12% Bis-Tris acrylamide NuPAGE gradient gels in MOPS SDS buffer (Invitrogen). The following antibodies were used: HSP90 (610419, BD), BRAF (sc5284, Santa Cruz), pERK (9101, Cell Signaling), and tERK (9102, Cell Signaling). Secondary HRP-tagged antibodies and ECL detection reagent were from Amersham. Image J software was used for quantification.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.043>.

### AUTHOR CONTRIBUTIONS

F.A.K. and P.P.P. conceived and designed the study. F.A.K., M.R., A.R., Y.T., D.W., N.S., and A.B. performed experiments, and F.A.K., M.R., and P.P.P. analyzed most data. C.N., J.F., and F.L. carried out immunohistochemistry. U.A. performed miRNA predictions. R.C. and S.J.R. evaluated histopathology of tumors and advised on immunohistochemical validation. M.S., T.M.K., M.D.C.V.-H., and D.J.A. performed RNA-seq analysis. B.C. and M.A.S. provided human samples and cell lines. A.V. and O.E. analyzed *BRAF*P1 expression in human RNA-seq data. F.A.K. and P.P.P. wrote the manuscript with contributions from all authors.

### ACKNOWLEDGMENTS

We are grateful to L. Cantley for hosting F.A.K. in his lab for part of this study. We thank S. Lowe for CAG-rTA3 mice; L. Cantley, B. Vogelstein, and P. Sharp for cell lines; G. DeNicola for helpful discussions; and the Nikon Imaging Cen-

ter at Harvard Medical School for help with light microscopy, the Dana Farber/Harvard Cancer Center Specialized Histopathology Core for help with immunohistochemistry, J. Clohessy and B. Padmani for help with ultrasound imaging, and S. Annunziato, A. Nyein, A. Bester, and K. Berry for technical assistance. F.A.K. was supported by fellowships from the Department of Defense Prostate Cancer Research Program and the American Cancer Society. M.R. was supported by the German Academy of Sciences Leopoldina, F.L. received a CHB-MIT fellowship, and U.A. acknowledges support from the Italian Association for Cancer Research (AIRC) under grant IG-9408. M.S., T.M.K., M.D.C.V.-H., and D.J.A. are funded by Cancer Research UK and the Wellcome Trust. R.C. was supported by a FP7 ERC-2009-StG grant (Proposal No. 242965, "Lunely"), AIRC grant IG-12023, and an International Association for Cancer Research (AICR) grant 12-0216. P.P.P. was supported by NIH grant CA170158-01.

Received: August 26, 2014

Revised: December 19, 2014

Accepted: February 2, 2015

Published: April 2, 2015

### REFERENCES

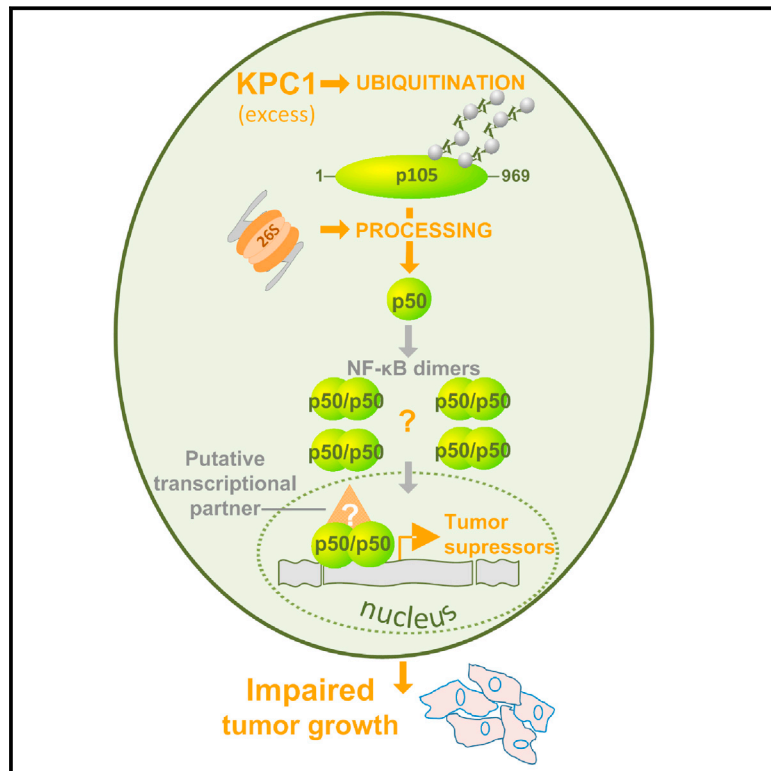
- Ala, U., Karreth, F.A., Bosia, C., Pagnani, A., Taulli, R., Léopold, V., Tay, Y., Provero, P., Zecchina, R., and Pandolfi, P.P. (2013). Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proc. Natl. Acad. Sci. USA* **110**, 7154–7159.
- Bea, S., Zettl, A., Wright, G., Salaverria, I., Jehn, P., Moreno, V., Burek, C., Ott, G., Puig, X., Yang, L., et al.; Lymphoma/Leukemia Molecular Profiling Project (2005). Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood* **106**, 3183–3190.
- Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23–28.
- Bier, A., Oviedo-Landaverde, I., Zhao, J., Mamane, Y., Kandouz, M., and Batist, G. (2009). Connexin43 pseudogene in breast cancer cells offers a novel therapeutic target. *Mol. Cancer Ther.* **8**, 786–793.
- Bosia, C., Pagnani, A., and Zecchina, R. (2013). Modelling Competing Endogenous RNA Networks. *PLoS ONE* **8**, e66609.
- Bosson, A.D., Zamudio, J.R., and Sharp, P.A. (2014). Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol. Cell* **56**, 347–359.
- Cazalla, D., Yario, T., and Steitz, J.A. (2010). Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science* **328**, 1563–1566.
- Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358–369.
- Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G.M., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790.
- Chiefari, E., Iiritano, S., Paonessa, F., Le Pera, I., Arcidiacono, B., Filocamo, M., Foti, D., Liebhauer, S.A., and Brunetti, A. (2010). Pseudogene-mediated posttranscriptional silencing of HMG1A can result in insulin resistance and type 2 diabetes. *Nat. Commun.* **1**, 40–47.
- Chiu, H.-S., Llobet-Navas, D., Yang, X., Chung, W.-J., Ambesi-Impimbato, A., Iyer, A., Kim, H.R., Seviour, E.G., Luo, Z., Sehgal, V., et al. (2014). Cupid: simultaneous reconstruction of microRNA-target and ceRNA networks. *Genome Res.* **25**, 257–267.
- Cummins, J.M., He, Y., Leary, R.J., Pagliarini, R., Diaz, L.A., Jr., Sjoblom, T., Barad, O., Bentwich, Z., Szafranska, A.E., Labourier, E., et al. (2006). The colorectal microRNAome. *Proc. Natl. Acad. Sci. USA* **103**, 3687–3692.



- Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell* 54, 766–776.
- Dewald, G.W., Brecher, M., Travis, L.B., and Stupca, P.J. (1989). Twenty-six patients with hematologic disorders and X chromosome abnormalities. Frequent idic(X)(q13) chromosomes and Xq13 anomalies associated with pathologic ringed sideroblasts. *Cancer Genet. Cytogenet.* 42, 173–185.
- Dierlamm, J., Michaux, L., Criel, A., Wlodarska, I., Zeller, W., Louwagie, A., Michaux, J.L., Mecucci, C., and Van den Berghe, H. (1995). Isodicentric (X)(q13) in haematological malignancies: presentation of five new cases, application of fluorescence in situ hybridization (FISH) and review of the literature. *Br. J. Haematol.* 91, 885–891.
- Figliuzzi, M., Marinari, E., and De Martino, A. (2013). MicroRNAs as a selective channel of communication between competing RNAs: a steady-state theory. *Biophys. J.* 104, 1203–1213.
- Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037.
- Gutschner, T., and Diederichs, S. (2012). The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.* 9, 703–719.
- Han, Y.J., Ma, S.F., Yourek, G., Park, Y.-D., and Garcia, J.G.N. (2011). A transcribed pseudogene of MYLK promotes cell proliferation. *FASEB J.* 25, 2305–2312.
- Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388.
- Hawkins, P.G., and Morris, K.V. (2010). Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. *Transcription* 1, 165–175.
- Heinonen, K., Mahlamäki, E., Riikonen, P., Meltoranta, R.L., Rahiala, J., and Perkiö, M. (1999). Acquired X-chromosome aneuploidy in children with acute lymphoblastic leukemia. *Med. Pediatr. Oncol.* 32, 360–365.
- Jens, M., and Rajewsky, N. (2014). Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat. Rev. Genet.* 16, 113–126.
- Johnsson, P., Ackley, A., Vidarsdottir, L., Lui, W.-O., Corcoran, M., Grandér, D., and Morris, K.V. (2013). A pseudogene long-noncoding-RNA network regulates PTEN transcription and translation in human cells. *Nat. Struct. Mol. Biol.* 20, 440–446.
- Kalyana-Sundaram, S., Kumar-Sinha, C., Shankar, S., Robinson, D.R., Wu, Y.-M., Cao, X., Asangani, I.A., Kothari, V., Prensner, J.R., Lonigro, R.J., et al. (2012). Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* 149, 1622–1634.
- Karreth, F.A., DeNicola, G.M., Winter, S.P., and Tuveson, D.A. (2009). C-Raf inhibits MAPK activation and transformation by B-Raf(V600E). *Mol. Cell* 36, 477–486.
- Karreth, F.A., Tay, Y., Perna, D., Ala, U., Tan, S.M., Rust, A.G., DeNicola, G., Webster, K.A., Weiss, D., Perez-Mancera, P.A., et al. (2011). In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell* 147, 382–395.
- Libri, V., Helwak, A., Miesen, P., Santhakumar, D., Borger, J.G., Kudla, G., Grey, F., Tollervey, D., and Buck, A.H. (2012). Murine cytomegalovirus encodes a miR-27 inhibitor disguised as a target. *Proc. Natl. Acad. Sci. USA* 109, 279–284.
- Marcinowski, L., Tanguy, M., Krmpotic, A., Rädle, B., Lisnić, V.J., Tuddenham, L., Chane-Woon-Ming, B., Ruzsics, Z., Erhard, F., Benkartek, C., et al. (2012). Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo. *PLoS Pathog.* 8, e1002510.
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338.
- Monni, O., Joensuu, H., Franssila, K., and Knuutila, S. (1996). DNA copy number changes in diffuse large B-cell lymphoma—comparative genomic hybridization study. *Blood* 87, 5269–5278.
- Morin, R.D., Mungall, K., Pleasance, E., Mungall, A.J., Goya, R., Huff, R.D., Scott, D.W., Ding, J., Roth, A., Chiu, R., et al. (2013). Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. *Blood* 122, 1256–1265.
- Paulsson, K., Haferlach, C., Fonatsch, C., Hagemeijer, A., Andersen, M.K., Slovak, M.L., and Johansson, B. (2010). The idic(X)(q13) in myeloid malignancies: breakpoint clustering in segmental duplications and association with TET2 mutations. *Hum. Mol. Genet.* 19, 1507–1514.
- Poliseno, L. (2012). Pseudogenes: newly discovered players in human cancer. *Sci. Signal.* 5, re5.
- Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., and Pandolfi, P.P. (2010). A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* 465, 1033–1038.
- Premisruti, P.K., Dow, L.E., Kim, S.Y., Camiolo, M., Malone, C.D., Miething, C., Scoppo, C., Zuber, J., Dickins, R.A., Kogan, S.C., et al. (2011). A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell* 145, 145–158.
- Rack, K.A., Chelly, J., Gibbons, R.J., Rider, S., Benjamin, D., Lafrenière, R.G., Oscier, D., Hendriks, R.W., Craig, I.W., Willard, H.F., et al. (1994). Absence of the XIST gene from late-replicating isodicentric X chromosomes in leukaemia. *Hum. Mol. Genet.* 3, 1053–1059.
- Ravi, A., Gurtan, A.M., Kumar, M.S., Bhutkar, A., Chin, C., Lu, V., Lees, J.A., Jacks, T., and Sharp, P.A. (2012). Proliferation and tumorigenesis of a murine sarcoma cell line in the absence of DICER1. *Cancer Cell* 21, 848–855.
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146, 353–358.
- Sumazin, P., Yang, X., Chiu, H.-S., Chung, W.-J., Iyer, A., Llobet-Navas, D., Rajbhandari, P., Bansal, M., Guarnieri, P., Silva, J., and Califano, A. (2011). An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell* 147, 370–381.
- Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., and Hannon, G.J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453, 534–538.
- Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F., et al. (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* 147, 344–357.
- Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev. Cell* 25, 69–80.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453, 539–543.
- Yamamoto, K., Nagata, K., Kida, A., and Hamaguchi, H. (2002). Acquired gain of an X chromosome as the sole abnormality in the blast crisis of chronic neutrophilic leukemia. *Cancer Genet. Cytogenet.* 134, 84–87.
- Zou, M., Baitei, E.Y., Alzahrani, A.S., Al-Mohanna, F., Farid, N.R., Meyer, B., and Shi, Y. (2009). Oncogenic activation of MAP kinase by BRAF pseudogene in thyroid tumors. *Neoplasia* 11, 57–65.

# KPC1-Mediated Ubiquitination and Proteasomal Processing of NF- $\kappa$ B1 p105 to p50 Restricts Tumor Growth

## Graphical Abstract



## Authors

Yelena Kravtsova-Ivantsiv,  
Inna Shomer, ..., Ilana Doweck,  
Aaron Ciechanover

## Correspondence

aaroncie@tx.technion.ac.il

## In Brief

KPC1-dependent ubiquitination of p105 supports limited processing to produce p50, thereby shaping the population of NF- $\kappa$ B subunits available. Production of p50 appears linked to a tumor-suppressive expression pattern and suggests a balancing of NF- $\kappa$ B's roles in tumorigenesis.

## Highlights

- KPC1 ubiquitinates p105, resulting in its basal and signal-induced cleavage to p50
- KPC1 inhibits tumor growth via regulation of p50-dependent tumor suppressor genes
- KPC1 and p50 levels are correlated and are lower in tumors than in normal tissue
- Excess p50 downregulates p65, possibly lowering tumorigenic levels of p50-p65 NF- $\kappa$ B

## Accession Numbers

GSE60530



# KPC1-Mediated Ubiquitination and Proteasomal Processing of NF- $\kappa$ B1 p105 to p50 Restricts Tumor Growth

Yelena Kravtsova-Ivantsiv,<sup>1,13</sup> Inna Shomer,<sup>1,13</sup> Victoria Cohen-Kaplan,<sup>1</sup> Berend Snijder,<sup>2</sup> Giulio Superti-Furga,<sup>2</sup> Hedva Gonen,<sup>1</sup> Thomas Sommer,<sup>3</sup> Tamar Ziv,<sup>4</sup> Arie Admon,<sup>4</sup> Inna Naroditsky,<sup>5</sup> Muhammad Jbara,<sup>6</sup> Ashraf Brik,<sup>6,7</sup> Eli Pikarsky,<sup>8,9</sup> Yong Tae Kwon,<sup>10,11</sup> Ilana Doweck,<sup>12</sup> and Aaron Ciechanover<sup>1,10,11,\*</sup>

<sup>1</sup>The David and Janet Polak Cancer and Vascular Biology Research Center, The Rappaport Faculty of Medicine and Research Institute, Technion-Israel Institute of Technology, Haifa 31096, Israel

<sup>2</sup>CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna 1090, Austria

<sup>3</sup>Max Delbrück Center for Molecular Medicine, Berlin 13125, Germany

<sup>4</sup>The Smoler Proteomics Center, Faculty of Biology, Technion-Israel Institute of Technology, Haifa 32000, Israel

<sup>5</sup>Department of Pathology, Rambam Health Care Campus, Haifa 31096, Israel

<sup>6</sup>Department of Chemistry, Ben-Gurion University of the Negev, Beer Sheva 84105, Israel

<sup>7</sup>The Schulich Faculty of Chemistry, Technion Israel Institute of Technology, Haifa 32000, Israel

<sup>8</sup>Department of Immunology and Cancer Research

<sup>9</sup>Department of Pathology

Institute for Medical Research Israel-Canada (IMRIC), Hebrew University-Hadassah Medical School, Jerusalem 91120, Israel

<sup>10</sup>Protein Metabolism Medical Research Center

<sup>11</sup>Department of Biomedical Sciences

College of Medicine, Seoul National University, Seoul 110-799, South Korea

<sup>12</sup>Department of Otolaryngology, Head and Neck Surgery, Carmel Medical Center, Haifa 34367, Israel

<sup>13</sup>Co-first author

\*Correspondence: [aaroncie@tx.technion.ac.il](mailto:aaroncie@tx.technion.ac.il)

<http://dx.doi.org/10.1016/j.cell.2015.03.001>

## SUMMARY

NF- $\kappa$ B is a key transcriptional regulator involved in inflammation and cell proliferation, survival, and transformation. Several key steps in its activation are mediated by the ubiquitin (Ub) system. One uncharacterized step is limited proteasomal processing of the NF- $\kappa$ B1 precursor p105 to the p50 active subunit. Here, we identify KPC1 as the Ub ligase (E3) that binds to the ankyrin repeats domain of p105, ubiquitinates it, and mediates its processing both under basal conditions and following signaling. Overexpression of KPC1 inhibits tumor growth likely mediated via excessive generation of p50. Also, overabundance of p50 downregulates p65, suggesting that a p50-p50 homodimer may modulate transcription in place of the tumorigenic p50-p65. Transcript analysis reveals increased expression of genes associated with tumor-suppressive signals. Overall, KPC1 regulation of NF- $\kappa$ B1 processing appears to constitute an important balancing step among the stimulatory and inhibitory activities of the transcription factor in cell growth control.

## INTRODUCTION

The NF- $\kappa$ B family of transcription factors is involved in regulation of a variety of genes that control the immune and inflammatory

response, cell survival and death, proliferation, and differentiation. Recently—150 years after Rudolf Virchow discovered the infiltration of tumors with leukocytes and proposed a linkage between chronic inflammation and malignant transformation—it has been shown that the mechanism(s) that underlies this linkage is mediated largely by the NF- $\kappa$ B family of transcription factors (Ben-Neriah and Karin, 2011; DiDonato et al., 2012). NF- $\kappa$ B is overexpressed in numerous tumors. It upregulates expression of anti-apoptotic genes such as IAPs, cell-cycle promoters, and growth factors and their receptors (DiDonato et al., 2012). Nevertheless, in some cases NF- $\kappa$ B was shown to display strong tumor-suppressive characteristics (Perkins, 2012; Pikarsky and Ben-Neriah, 2006). For example, it is involved in regulation of activation-induced apoptosis of T lymphocytes (Ivanov et al., 1997) and in inducing cell-cycle arrest and cell death caused by repression of Bcl2, XIAP, Bcl-X<sub>L</sub>, Cyclin D1, and c-Myc that occurs after cell damage. The arrest and death are mediated by p52 dimers (Barré et al., 2010; Barré and Perkins, 2007). Also, it was shown that NF- $\kappa$ B1<sup>-/-</sup> cells accumulate alkylator-induced mutations, and NF- $\kappa$ B1<sup>-/-</sup> mice develop more lymphomas following alkylating agent-induced DNA damage, again suggesting that NF- $\kappa$ B1 can act as a tumor suppressor (Voce et al., 2014).

The family members are mostly heterodimers where one of the subunits—p52 or p50—is the product of limited, ubiquitin- and proteasome-mediated processing of a longer (and inactive) precursor, p100 or p105, respectively (Betts and Nabel, 1996; Fan and Maniatis, 1991; Palombella et al., 1994). The other subunit is typically a member of the Rel family of proteins (RelA-p65, RelB, or c-Rel). At times, p50 and p52 can generate homodimers

that cannot act as transcriptional activators since they lack a transactivation domain present in the Rel proteins. In unstimulated cells, the NF- $\kappa$ B dimers are sequestered in the cytosol attached to ankyrin repeats (ARs) of I $\kappa$ B inhibitory proteins (I $\kappa$ B, Bcl3, p100, and p105). A broad array of extracellular signals stimulate degradation of the I $\kappa$ B proteins, resulting in translocation of the dimers to the nucleus where they initiate different transcriptional programs (Rahman and McFadden, 2011).

Proteasomal processing of p105 occurs under both basal conditions and following stimulation and requires prior ubiquitination (Cohen et al., 2004; MacKichan et al., 1996). One element that was shown to be important in the processing is a long Gly-Ala repeat in the middle of p105 that may serve as a proteasomal “stop signal” (Lin and Ghosh, 1996). In addition to processing, p105 can also undergo complete degradation, releasing NF- $\kappa$ B dimers anchored to its C-terminal ARs domain. Following stimulation, p105 is phosphorylated on serine residues 927 and 932 by I $\kappa$ B kinase (IKK $\beta$ ) (Salmerón et al., 2001). This modification recruits the beta-Transducin Repeat Containing Protein ( $\beta$ TrCP) E3 (Oran et al., 2000), resulting in complete degradation of the molecule (Heissmeyer et al., 2001). The ligase(s) involved in processing of p105 under basal conditions as well as following stimulation has remained elusive.

In the present study, we identified KIP1 ubiquitination-promoting complex (KPC) as the Ub ligase that is involved in both basal and signal-induced processing of p105. KPC is a heterodimer made of KPC1 (RNF123) and KPC2 (UBAC1). It was shown to degrade the cyclin-dependent kinase inhibitor p27Kip1 in the G1 phase of the cell cycle (Kamura et al., 2004). KPC1 is a RING-finger protein that serves as the ligase. KPC2 interacts with ubiquitinated proteins and with the proteasome via its two Ub-associated domains and a Ub-like domain, acting as a shuttle that promotes the degradation of p27Kip1. It was also shown to stabilize KPC1 (Hara et al., 2005).

## RESULTS

### Identification of KPC1 as the p105 Ub Ligase

One of the still missing links in the Ub-mediated activation pathway of NF- $\kappa$ B is the identity of the ligase that ubiquitinates p105, resulting in its proteasomal processing to the p50 active subunit. To identify the ligase, we sequentially fractionated rabbit reticulocyte lysate using different chromatographic principles (Figure 1Ai). Each fraction along the different steps was monitored for E3 activity in a cell-free reconstituted conjugation assay containing in vitro translated <sup>35</sup>S-labeled p105 as a substrate (Figure 1Aii). To avoid ubiquitination by the  $\beta$ TrCP ligase, we used p105S927A mutant that cannot be phosphorylated by IKK $\beta$  and therefore cannot bind this E3. Employing mass spectrometric analysis, peptides derived from the KPC Ub ligase were identified in active fractions along the three last chromatographic steps. In the last step of purification (heparin), we identified 58 KPC1 peptides and seven KPC2 peptides covering 43.21% and 19.8% of the open reading frames, respectively (Figure 1B). Because of lack sequence information on rabbit KPC2, we used the sequence of the mouse protein to demonstrate the coverage map. The changes between the two species are negligible (but shown).

To test directly the role of KPC in p105 ubiquitination and processing, we established a cell-free conjugation assay using labeled p105 as a substrate and purified KPC1 or its catalytically inactive species (mutated in the RING domain) KPC1I1256A as the ligase. The wild-type (WT) ligase catalyzed conjugation of p105, whereas the inactive ligase did not (Figure 2A). It appears that KPC1 activity is specific to p105, as it scarcely modifies p100 that is highly homologous to p105 and also undergoes limited proteasomal processing, most probably by a different ligase (Figure S1A).

To demonstrate the ability of KPC1 to modify p105 in cells, we overexpressed Flag-p105 along with HA-Ub in HEK293 cells, in which KPC1 was either silenced (Figure 2B, lane 1), or overexpressed (Figure 2B, lanes 2 and 3). Immunoprecipitation of p105 revealed that it is sparsely ubiquitinated in the absence of the ligase, and ubiquitination is increased significantly following overexpression of KPC1 (Figure 2Bi; immunoprecipitation [IP], compare lanes 1 and 2). Furthermore, we found that p105 binds to KPC1 and co-immunoprecipitates with it (Figure 2Biii; IP, lane 2). In addition, we demonstrated that endogenous KPC1 interacts with endogenous p105 (Figure S1B).

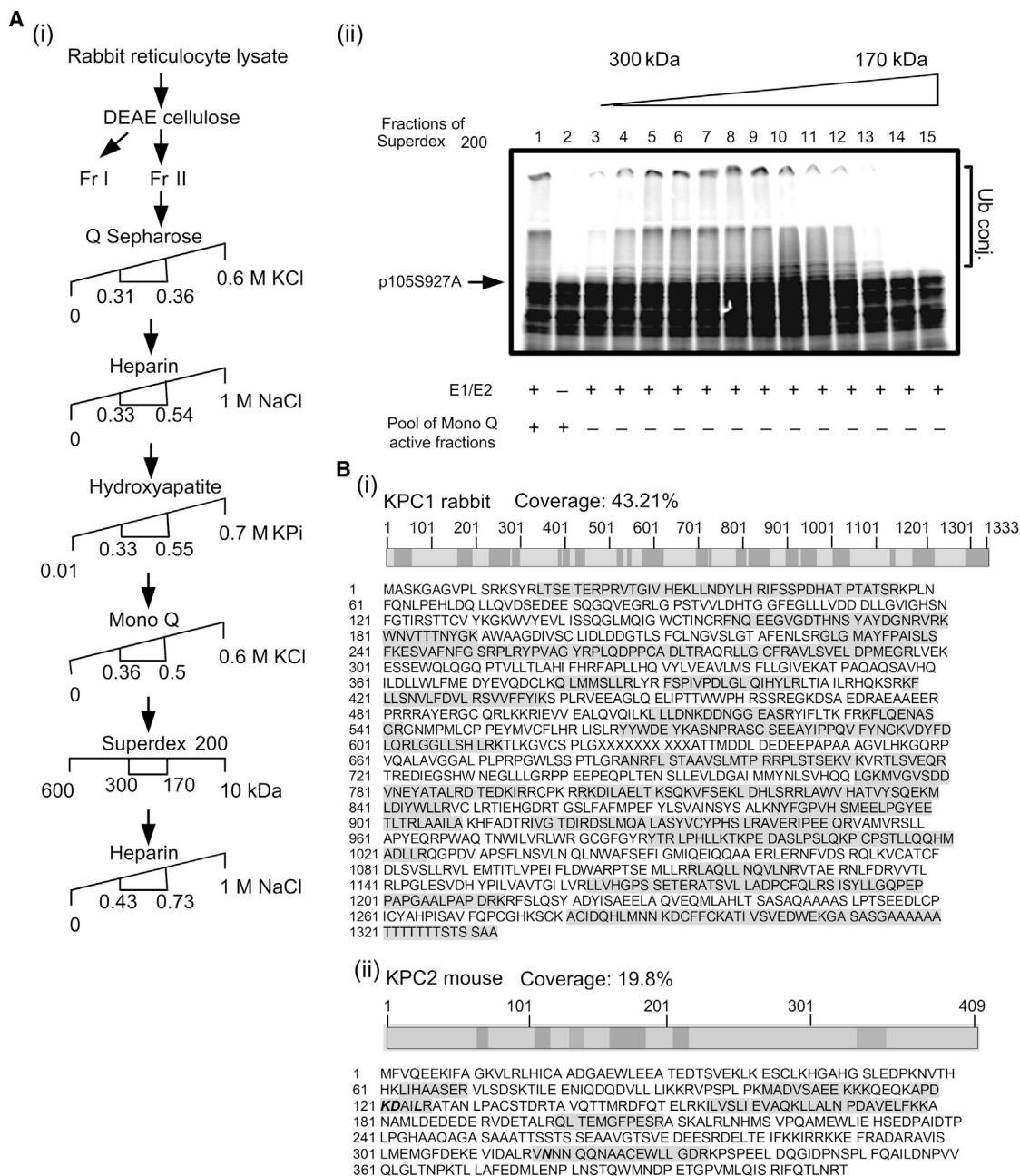
### KPC1 Promotes Basal and Signal-Induced Processing of p105

To demonstrate the involvement of KPC1 in p105 processing, we silenced its expression in cells using small interfering RNA (siRNA). As can be seen in Figure 2C, the silencing of KPC1 decreased the amount of p50 generated from p105. In a different experiment, we expressed in HEK293 cells FLAG-p105 along with Myc-KPC1 or Myc-KPC1I1256A. Less p50 was generated in the presence of the KPC1 mutant (Figure S1C).

As noted, processing of p105 occurs also following stimulation. It was interesting to study whether KPC1 can promote p105 processing under these conditions as well. Therefore, we tested the generation of p50 from p105 following expression of constitutively active IKK $\beta$  (IKK $\beta$ S176,180E) in the presence (endogenous) or absence (silenced) of KPC1. As expected, the stimulation increased the processing of p105 (compare Figure 2D to Figure 2C; control siRNA). Silencing of KPC1 significantly decreased the generation of p50 following stimulation, strongly suggesting a role for KPC1 in signal-induced processing (Figure 2D). It is known that under the influence of the kinase, the precursor was not only processed but also degraded to a significant extent (compare Figure 2D to Figure 2C and note in particular the decreasing amount of p105 + p50 remained along time following stimulation). It should be noted that the degradation rate of p105 following stimulation was significantly higher in cells that lack KPC1 (Figure 2D). It is possible that the processing of p105 mediated by KPC1 and its degradation mediated by  $\beta$ TrCP occur in parallel. When one process is inactivated, the other becomes dominant. The influence of KPC1 on signal induced-processing of p105 appears to be specific, as its silencing does not affect the processing of p100 following NF- $\kappa$ B-inducing kinase (NIK) expression (Figure S1D).

In all these experiments, we used exogenously expressed p105. To demonstrate the effect on endogenous p105, we used the human haploid cell line HAP1 in which the single allele of KPC1 or KPC2 were knocked out using the Crispr-CAS





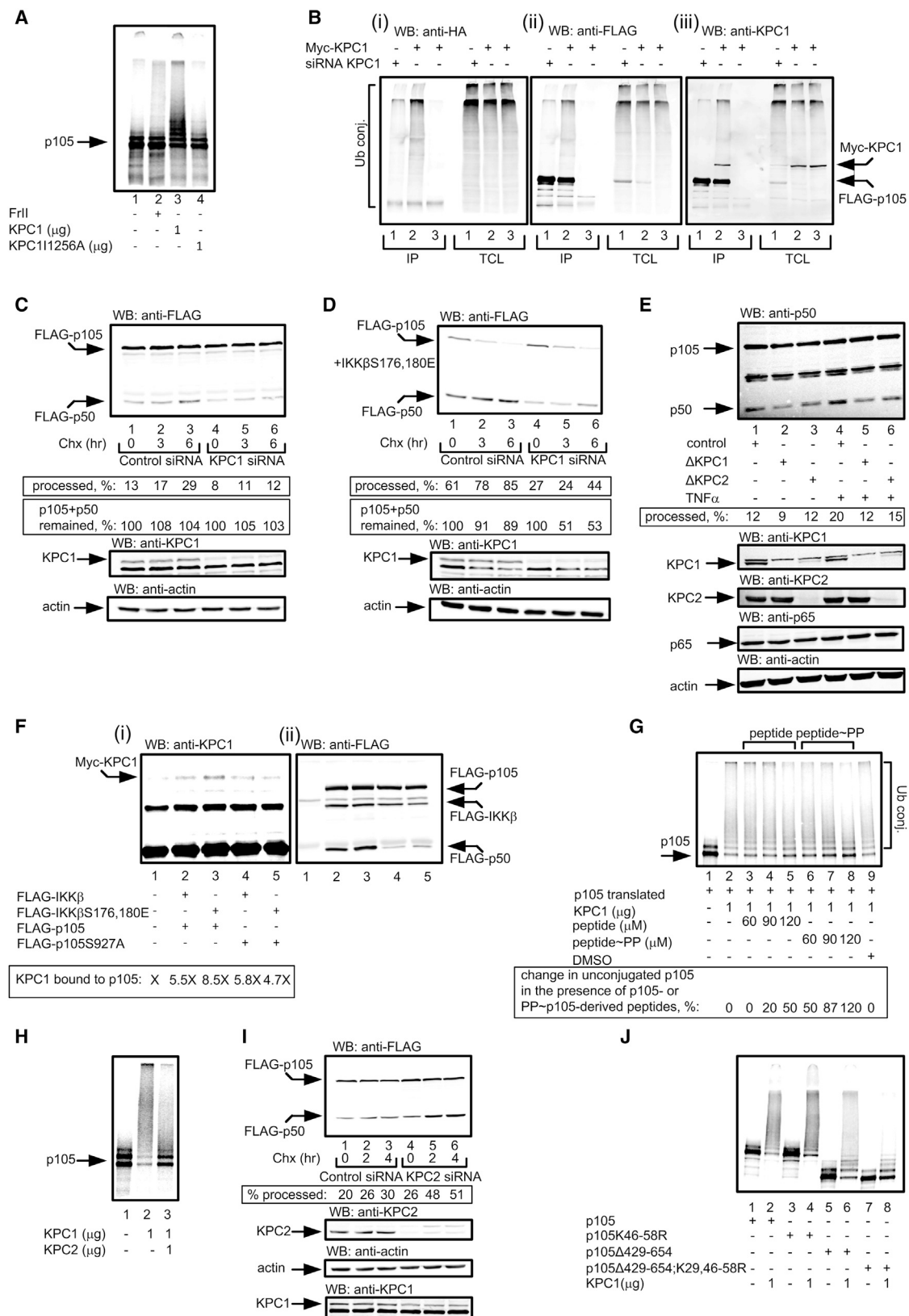
**Figure 1. Purification and Identification of the p105 Ub Ligase**

(A) (i) Scheme of the chromatographic resolution of Fraction II monitoring the E3 ligating activity toward p105. Numbers represent salt concentrations (M) or molecular weight (kDa) at which the ligating activity was eluted from the respective columns. Fr II, Fraction II. (ii) E3 conjugating activity profile along the fractions resolved by the Superdex 200 gel filtration column. In vitro translated and  $^{35}\text{S}$ -labeled p105S927A was ubiquitinated in a reconstituted cell-free system in the presence of the resolved fractions.

(B) (i) Peptide coverage map of rabbit KPC1. The peptides were identified through mass spectrometric analysis of the E3-containing fractions resolved by the last, Heparin-based column. (ii) Peptide coverage map of mouse KPC2. The peptides were identified through mass spectrometric analysis of the E3-containing fractions resolved by the last, Heparin-based column. Residues marked in bold and italics denote differences in sequence between mouse and rabbit.

technology. Elimination of KPC1 or KPC2 (that stabilizes KPC1 [Hara et al., 2005]; note that removal of KPC2 results in a significant decrease in the level of KPC1; Figure 2E) decreased the generation of p50 both in the presence or absence of  $\text{TNF}\alpha$

(Figure 2E). In contrast, the level of p65 was not affected. The finding that p50 is still present, albeit in a decreased level, in the KPC1 KO cells, may be due to the activity of another, yet to be identified ligase, and/or to co-translational processing of



(legend on next page)

the nascent peptide that occurs before completion of the p105 precursor synthesis (Lin et al., 1998). It should also be noted that the effect of KPC1 on p50 generation is significantly more pronounced in tumors growing in mice than in cultured cells (see below).

Our finding that KPC1 mediates processing under both basal and stimulated conditions prompted us to dissect the mechanism involved. We monitored the interaction between KPC1 and p105 under basal and stimulated conditions and found that expression of constitutively active IKK $\beta$  results in increased interaction between the two as assayed by co-immunoprecipitation (Figures 2F and S1E). The finding that the interaction of p105S927A with KPC1 is not affected by IKK $\beta$  (Figure 2Fi, lanes 4 and 5) attests to the specificity of the effect of IKK $\beta$  in phosphorylating a specific Ser residue (927) in p105. As expected, we found that ubiquitination of phosphorylated p105 by KPC1 is stronger compared to that of the non-phosphorylated species (Figure S1F).

To further confirm that KPC1 interacts more efficiently with phosphorylated p105, we designed an experiment in which we competed on the binding of p105 to the ligase with a synthetic phosphorylated peptide derived from the p105 IKK $\beta$ -phosphorylation site. The phosphorylated peptide inhibited ubiquitination of p105 by KPC1 to a larger extent compared with its

non-phosphorylated species, both in a crude system and in a system made of purified components (Figures 2G and S1G, respectively).

### Role of KPC2 in KPC1-Mediated p105 Ubiquitination and Processing

At that point, it was important to study the role of KPC2, the partner of KPC1 in the heterodimeric ligase complex, in p105 modification and processing. We noted that its addition to a reconstituted cell-free system decreases significantly the ubiquitination of p105 by KPC1 (Figure 2H). This was true also when p105 was purified by a specific antibody, ruling out a possible effect of other components present in the mixture in which the labeled p105 was translated (Figure S2Ai). To rule out that the reduced ubiquitination of p105 in the presence of KPC2 is due to a possible deubiquitinating activity of the protein, we added it to the cell-free ubiquitination system after KPC1, when most of the ubiquitination reaction was completed. It had no effect on the conjugates pattern (Figure S2Aii). The interference of KPC2 in chain formation appears to be specific to KPC1 and p105, as it did not affect the ligase activity of E6-AP toward RING1B<sup>153S</sup> (Zaaroor-Regev et al., 2010) (Figure S2B).

Importantly, in correlation with the suppressive effect of KPC2 on KPC1-mediated ubiquitination of p105, silencing of KPC2

### Figure 2. p105 Is a Substrate of KPC1 in a Cell-free System and in Cells, Both under Basal Conditions and following Signaling

(A) Ubiquitination of in vitro translated and <sup>35</sup>S-labeled p105 by Fraction II and purified KPC1-FLAG-TEV-6xHIS or KPC111256A-FLAG-TEV-6xHIS in a reconstituted cell-free system. Fr II, Fraction II.

(B) KPC1 ubiquitinates p105 in cells. HEK293 cells that were transfected with siRNA to silence KPC1 (lane 1) or with control siRNA (lanes 2 and 3), were also transfected with cDNAs coding for FLAG-p105 (lanes 1 and 2), HA-Ub (lanes 1–3), and Myc-KPC1 (lanes 2 and 3). FLAG-p105 and its conjugates were immunoprecipitated from the cell lysates using immobilized anti-FLAG (IP; lanes 1–3), resolved via SDS-PAGE, and visualized using anti-HA (Bi) or anti-FLAG (Bii). KPC1 was visualized using a specific antibody to the protein (Biii). Ten percent of total cell lysates (TCL; lanes 1–3) were analyzed for expression of FLAG-p105, HA-Ub or Myc-KPC1, using anti-HA (Bi), anti-FLAG (Bii), or anti-KPC1 (Biii), respectively. IP, immunoprecipitation; WB, western blot.

(C) Silencing of KPC1 affects basal processing of p105. HEK293 cells were transfected with control siRNA (lanes 1–3) or siRNA to silence KPC1 (lanes 4–6). After 24 hr, cells were transfected with cDNAs coding for FLAG-p105. Processing of p105 was calculated as the ratio between the amount p50 at the specified time and the sum of p50 + p105 at time zero (in order to disregard degradation of p105 in our calculations), multiplied by 100%. The amount of p50 + p105 remained (reflecting degradation along time) was calculated as the sum of p50 + p105 measured at the relevant time point, divided by the sum of p50 + p105 at time zero, multiplied by 100%.

(D) Silencing of KPC1 inhibits signal-induced processing of p105. HEK293 cells were transfected with control siRNA (lanes 1–3) or siRNA that targets KPC1 (lanes 4–6). After 24 hr, cells were transfected with cDNAs coding for FLAG-p105 and IKK $\beta$ S176,180E. Twenty-four hours after transfection (in the experiments depicted under C and D), cycloheximide was added for the indicated times, and cells were lysed, resolved via SDS-PAGE, and proteins visualized using anti-FLAG, anti-KPC1 or anti-actin. Processing and degradation were assessed as described under (C). Chx, cycloheximide. Actin was used to ascertain equal protein loading.

(E) Deletion of KPC1 or KPC2 genes inhibits basal and TNF $\alpha$ -induced processing of endogenous p105. Lysates were prepared from HAP1 control or HAP1 cells knocked out for the genes coding for KPC1 or KPC2. The lysates were resolved via SDS-PAGE, and proteins were visualized using anti-NF- $\kappa$ B1, anti-KPC1, anti-KPC2, anti-p65, or anti-actin. The amount of p105 processed was calculated as the ratio between the generated p50 and the sum of p50 + p105, multiplied by 100%.

(F) The interaction between p105 and KPC1 increases following signaling. HEK293 cells were transfected with cDNAs coding for FLAG-p105 (lanes 2 and 3) or FLAG-p105S927A (lanes 4 and 5) along with Myc-KPC1 (lanes 1–5) and FLAG-IKK $\beta$  (lanes 2 and 4) or FLAG-IKK $\beta$ S176,180E (lanes 3 and 5). FLAG-p105 and FLAG-p105S927A were immunoprecipitated from the cell lysate using immobilized anti-FLAG (lanes 1–5), and the bound KPC1 was visualized with anti-KPC1 (Fi). Immunoprecipitated p105s were visualized using anti-FLAG (Fii).

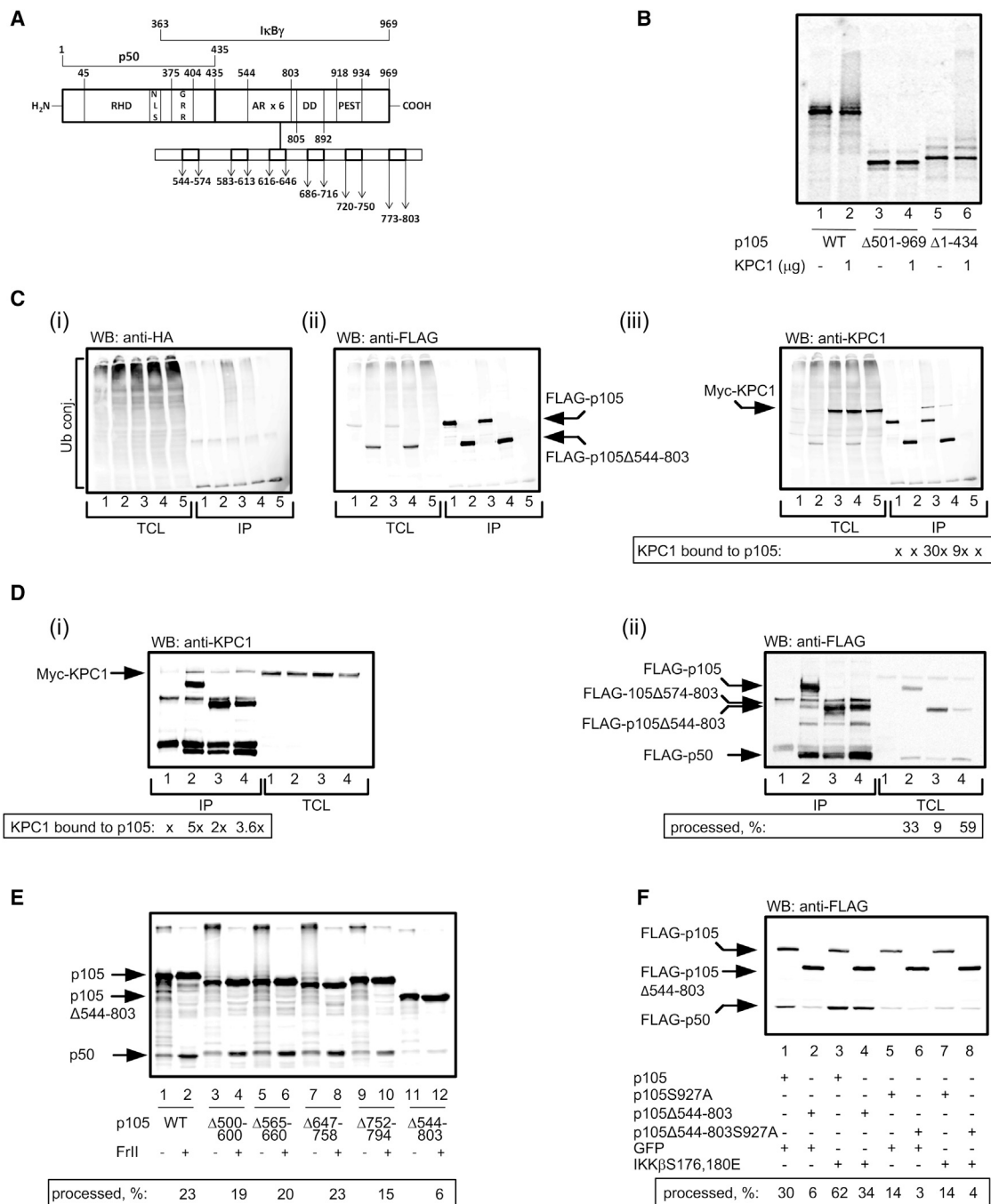
(G) A phosphorylated peptide corresponding to the signaled sequence in p105 inhibits its ubiquitination. In vitro translated and <sup>35</sup>S-labeled p105 was ubiquitinated by purified KPC1-FLAG-TEV-6xHIS (lanes 2–9) in a reconstituted cell-free system in the presence of a phosphorylated peptide derived from the signaled sequence of p105 (lanes 6–8), or in the presence of its non-phosphorylated counterpart (lanes 3–5). Presented is the change (in %) of unconjugated p105 remained following addition of increasing concentrations of the peptides (compared to a system to which a peptide was not added; lane 2).

(H) KPC2 attenuates ubiquitination of p105 by KPC1. Ubiquitination of in vitro translated and <sup>35</sup>S-labeled p105 by purified KPC1-FLAG-TEV-6xHIS in the presence or absence of HIS-KPC2 was carried out in a cell-free reconstituted system.

(I) KPC2 attenuates processing of p105 in cells. HEK293 cells were transfected with control siRNA (lanes 1–3) or siRNA to silence KPC2 (lanes 4–6). After 24 hr, cells were transfected with cDNAs coding for FLAG-p105 and generation of p50 was monitored 24 hr later. Processing of p105 was calculated as described under (C).

(J) KPC1 modifies lysine residues in the C-terminal segment of p105. In vitro-translated and <sup>35</sup>S-labeled WT and the indicated p105 mutants were subjected to ubiquitination by purified KPC1-FLAG-TEV-6xHIS in a reconstituted cell-free system.

See also Figures S1 and S2.



**Figure 3. KPC1-Dependent Ubiquitination and Processing of p105 Require the ARs of p105**

(A) Schematic representation of p105 domains. Numbers denote the respective residue along the protein sequence. RHD, Rel homology domain; NLS, nuclear localization signal; GRR, glycine rich repeat; AR, ankyrin repeats (all six of them are marked).

(B) The ARs-containing C-terminal half of p105 is ubiquitinated by KPC1. In vitro-translated and <sup>35</sup>S-labeled p105, p105Δ501-969 or p105Δ1-434 were subjected to ubiquitination by purified KPC1-FLAG-TEV-6xHis in a reconstituted cell-free system.

(C) The ARs of p105 are essential for binding of KPC1 and for its ubiquitination by the ligase in cells. HEK293 cells that were transfected with siRNA to silence KPC1 (lanes 1 and 2) or with control siRNA (lanes 3–5), were also transfected with cDNAs coding for FLAG-p105 (lanes 1 and 3), p105Δ544-803 (lanes 2 and 4), HA-Ub (lanes 1–5), and Myc-KPC1 (lanes 3–5). The different FLAG-p105 species and their conjugates were immunoprecipitated from the cell lysates by immobilized anti-FLAG (IP; lanes 1–5).

(D) KPC1 interacts with a single AR in p105. HEK293 cells were transfected with cDNAs coding for FLAG-p105 (lane 2), FLAG-p105Δ544-803 (lane 3), or FLAG-p105Δ574-803 (lane 4), along with Myc-KPC1 (lanes 1–4). The different FLAG-p105 species were immunoprecipitated from the cell lysates using immobilized anti-FLAG (IP; lanes 1–4).

(legend continued on next page)



increased the formation of p50 (Figure 2I). That, despite the fact that the short-term silencing reduced partially the level of KPC1 via its effect (or absence thereof) on the stabilization of the ligase (Figure 2I; note the change in the level of KPC1 following KPC2 silencing).

### Identification of the Ub Anchoring Sites on p105 Modified by KPC1

We have already shown that multiple lysines in the C-terminal segment of p105 are required for its ubiquitination and processing (Cohen et al., 2004; Kravtsova-Ivantsiv et al., 2009) in crude extracts. It was therefore important to show that this is true also for KPC1. Progressive removal of all lysine residues from the C-terminal segment (Figure S2C) resulted in corollary decrease in conjugation of p105 by KPC1 in a cell-free assay (Figure 2J) and in processing of the precursor in cells (Figure S2D).

### The C-Terminal ARs of p105 Are Necessary for Its Interaction with KPC1 and for Its Subsequent Ubiquitination and Processing

p105 harbors several domains: REL homology domain (RHD), nuclear localization signal (NLS), and a glycine rich repeat (GRR) in its N-terminal segment, and ARs, death domain (DD), and a PEST (proline, glutamate, serine, and threonine) sequence in the C-terminal segment (Figure 3A). We examined which of these domains is necessary for ubiquitination by KPC1. As can be seen in Figure 3B, removal of the C-terminal segment abolished altogether conjugation in a cell-free system, whereas removal of the N-terminal segment had no effect. Subsequently we found that removal of all six ARs (p105 $\Delta$ 544-803) affected significantly the ubiquitination of p105 by KPC1 (Figure S3A, compare lanes 2 and 12). Partial deletion of the repeats affected conjugation only slightly (compare lane 2 to lanes 4, 6, 8, and 10).

Similar results were obtained in experiments carried out in cells. Overexpression of KPC1 increased the ubiquitination of WT p105, but much less so of p105 that lacks all its ARs (Figure 3Ci; IP, compare lane 4 to lane 3). Importantly, in parallel, we observed also a decrease in the interaction between the ARs'-truncated p105 and its ligase compared to WT p105 (Figure 3Ciii; IP, compare lanes 4 and 3).

To rule out the possibility that the decrease in ubiquitination of p105 that lacks all its ARs is due to removal of the eight lysine residues in the repeats, we generated a mutant p105 in which all these lysines were substituted by arginines. The ubiquitination of the K to R mutant as well as its interaction with KPC1, were similar to that of WT p105 (Figures S3Bi and ii, respectively). An interesting question relates to the number of ARs necessary for ubiquitination and processing of p105. We constructed a

p105 mutant where all ARs except one have been deleted (p105 $\Delta$ 574-803). The single remaining AR was sufficient to bind KPC1 and to promote processing similar to that observed for WT p105 (Figure 3D). Thus, it appears that the ARs are redundant with relation to binding of KPC1.

Last, it was important to demonstrate whether the ARs-dependent ubiquitination increases the processing of p105. As can be seen in Figure 3E, mutant p105 that lacks all ARs, is processed much less efficiently compared to the WT species and to one lacking only some of the repeats (compare lane 12 to lanes 2, 4, 6, 8, and 10). A similar result was obtained also in cells (Figure 3F, lanes 1 and 2). Mutant p105 in which all lysine residues in the ARs were substituted with arginines (FLAG-p105K8R), is processed similarly to WT p105 (Figure S3C, lane 3), strongly suggesting that the ARs are required for the binding, ubiquitination, and processing of p105, but do not serve as ubiquitination sites essential for processing.

It appears that the ARs are also involved in signal-induced processing of p105, as their removal significantly decreased IKK $\beta$ -mediated generation of p50 (Figure 3F, compare lane 4 to lane 3). As expected, FLAG-p105S927A and FLAG-p105S927A $\Delta$ 544-803 did not respond to IKK $\beta$ -mediated phosphorylation (Figure 3F, lanes 7 and 8).

### Overexpression of KPC1 or p50 Suppresses Tumor Growth

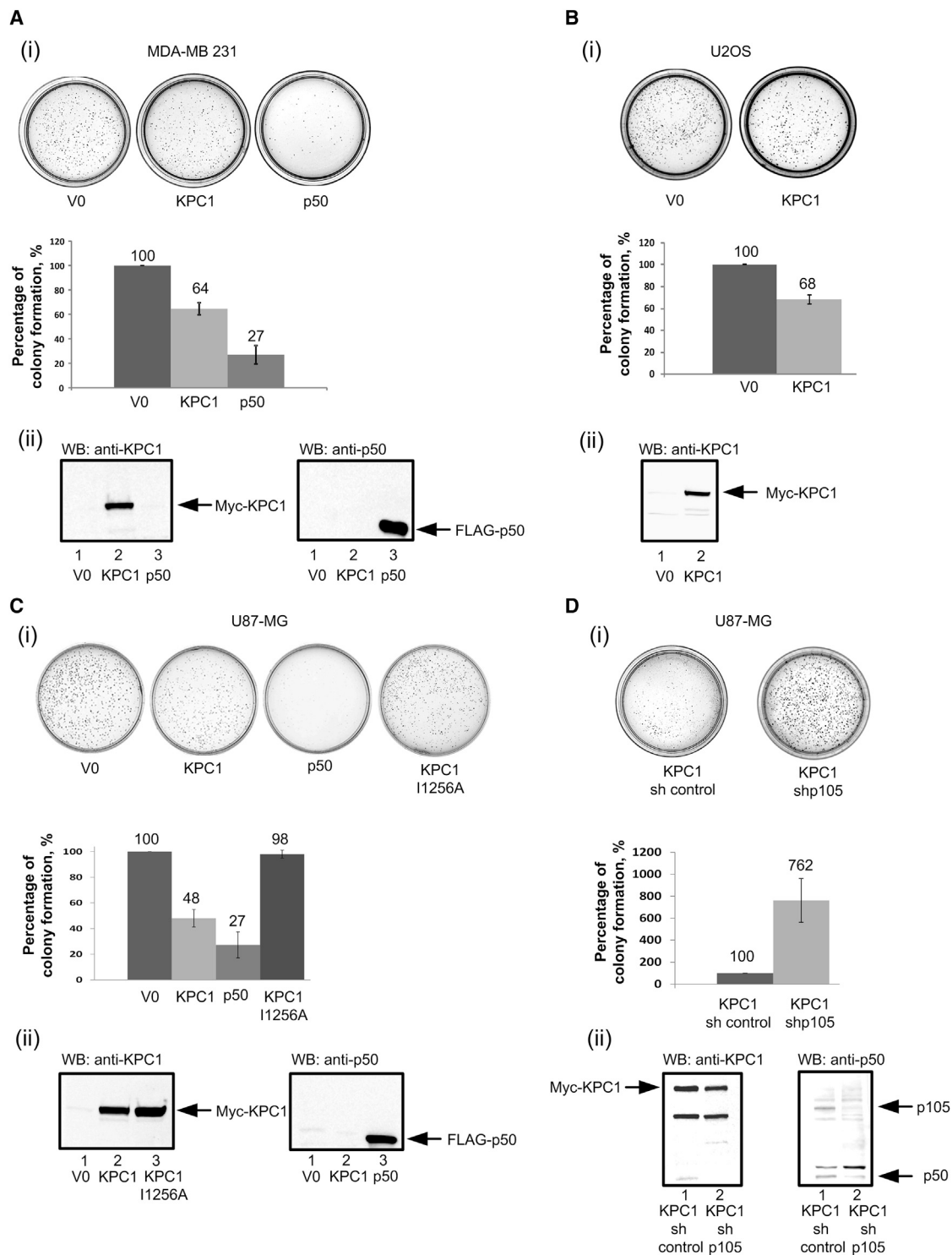
Since NF- $\kappa$ B dimers are known to affect cell survival, proliferation, and tumor progression, it was interesting to study the outcome of KPC1 on cell growth. Initially, we monitored the influence of overexpressed KPC1 on anchorage-independent growth in MB-MDA 231, U2OS, and U87-MG cells, and found that it inhibits colony formation by 36%, 32%, and 52%, respectively, compared to controls (Figures 4A–4C). Importantly, this effect was abrogated in cells overexpressing the inactive ligase species KPC1I1256A, suggesting that the inhibitory effect is due to the ligase activity (Figure 4C). Cells expressing p50 showed an even stronger inhibition of colony formation (73% for both MB-MDA 231 and U87-MG cells; Figures 4A and 4C), strongly suggesting that the effect of the ligase is mediated through its activity on p105, resulting in excessive generation of p50. Supporting the linkage is the finding that silencing of p105 abrogated the strong suppressive effect of KPC1: the number of colonies formed using cells that overexpress KPC1 in the absence of p105 was 7.5-fold larger than that formed in its presence (Figure 4D). It was important to study whether the growth suppressive effect of KPC1 and p50 is not due to induction of apoptosis. Thus, we stained U87-MG cells that overexpress these proteins for cleaved caspase 3. As can be seen in Figure S4, we could not detect the apoptotic marker. For that

(E) p105 that lacks its ARs is processed less efficiently in a cell-free system. The different <sup>35</sup>S-labeled p105 species were processed in a cell-free reconstituted system in the presence or absence of Fraction II as indicated. Fr II, Fraction II.

(F) Deletion of the ARs of p105 affects both its basal and signal-induced processing. HEK293 cells were transfected with cDNAs coding for FLAG-p105, FLAG-p105 $\Delta$ 544-803, FLAG-p105S927A, or FLAG-p105S927A $\Delta$ 544-803 along with either GFP or IKK $\beta$  as indicated.

In (C), (D), and (F), proteins were resolved via SDS-PAGE, blotted onto nitrocellulose membrane, and p105 and p50 were detected using anti-FLAG, KPC1 was detected using anti-KPC1, and Ub conjugates were detected using anti-HA. Ten percent of the total cell lysates (TCL) were analyzed for the expression of proteins. The SDS-PAGE-resolved labeled proteins in the experiments shown in (B) and (E) were visualized using PhosphorImaging. Processing was assessed as described under Figure 2E.

See also Figure S3.



**Figure 4. KPC1 and p50 Suppress Anchorage-Independent Growth of Cells**

Suppression of colony formation by overexpressed KPC1 or p50 in MDA-MB 231 (Ai), U2OS (Bi), and U87-MG (Ci and Di) cells. Cells were stably transfected with V0, or with cDNAs that code for Myc-KPC1, Myc-KPC1I1256A or FLAG-p50, or with cDNA coding for Myc-KPC1 along with control shRNA or shRNA to silence p105, as indicated, and were seeded on soft agar plates. After 3 weeks, the colonies were stained with 0.05% crystal violet. Data derived from five experiments ( $\pm$ SEM) are presented graphically. Expression of KPC1, KPC1I1256A, p50, and p105 is shown in (Aii), (Bii), (Cii), and (Dii).

See also [Figure S4](#).

experiment, it was also important to demonstrate that the suppressive effect of KPC1 and p50 is not due to some non-specific effect of the overexpression of the proteins. The unleashing of growth in the presence of overexpressed KPC1 but in the absence of p105 (Figure 4D), strongly suggests that the effect of KPC1 is indeed specific.

These observations prompted us to study the effect of KPC1 in a tumor model in mice. We generated xenografts stably overexpressing V0, KPC1, KPC1I1256A, or p50. Both the growth rate and weights of tumors expressing KPC1 and p50 were significantly smaller compared to those that harbor V0 or KPC1I1256A (Figures 5Ai and 5Aiii for xenografts derived from U87-MG cells, and Figure 5Aii for xenografts derived from MDA-MB 231 cells). Importantly, in tumors that overexpress KPC1, the level of p50 is significantly higher compared with tumors that express V0 (Figure 5Aiv), suggesting again a direct linkage between the KPC1 ligase activity and increased generation of p50. Interestingly, in tumors that overexpress KPC1 or p50, we also observed a significant decrease in the level of p65 (Figure 5Aiv). This finding raises the possibility that a different NF- $\kappa$ B transcription factor is generated under the influence of KPC1, possibly a p50 homodimer (see Discussion). To demonstrate that there are indeed changes in NF- $\kappa$ B species in human tumor xenografts overexpressing KPC1 and p50, we used electrophoretic mobility shift assay (EMSA) to monitor the activity of the transcription factor. As can be seen in Figure S5A, there is a significant decrease in the ability of “canonical” NF- $\kappa$ B to bind its consensus DNA sequence following overexpression of KPC1 and even more so following overexpression of p50.

Of note, all the effects on tumor growth (reduction in colony formation, tumor growth rate, and weight) were more prominent in p50-expressing tumors than in their KPC1-overexpressing counterparts. This is not surprising, as p50 is the product of KPC1 activity, and its direct expression has a stronger effect.

The functional linkage between KPC1 and p50 can also be observed in staining of specific proliferation and differentiation markers in the mice-derived tumors. The overexpression of KPC1, but not of KPC1I1256A, results in increased appearance of nuclear NF- $\kappa$ B (Figure 5B), a significant decrease in the proliferation marker ki-67, and an increase in the glial fibrillary acidic protein (GFAP), a known glial cells differentiation marker. Suspecting that KPC1 stimulates apoptosis, we looked for an increase in cleaved caspase 3, however, there was no change in the levels of the active enzyme compared to control sections. Staining of p27Kip1—a suggested substrate of KPC1 (Kamura et al., 2004)—did not show any change in the protein level (Figure 5B). This may be due to the differences in the systems studied.

### KPC1 Regulates Expression of a Subset of p50 Target Genes

We next analyzed the profile of gene expression in the tumors using RNA sequencing (RNA-seq) analysis of transcripts mapped to the human genome (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60530>; Table S1). The altered gene expression patterns revealed a strong similarity between overexpression of p50 and KPC1 in U87-MG xenografts (correlation of 0.51,  $p$  value  $< 10^{-300}$ ; Figure 6Ai), with 48 downregulated

and 534 upregulated genes that were consistent and significant in all replicates (Figure 6Aii; Table S2). The relative transcript levels of selected genes that were shown to be significantly upregulated and downregulated in RNA-seq analyses, was corroborated by quantitative real-time PCR (qRT-PCR) (Figure S5B). Functional analysis revealed highly significant enrichment in glycosylated and extracellular matrix proteins, and upregulation of genes expressing proteins involved in cell-cell and cell-substrate adhesion, regulation of cell migration, cell junctions, vasculature development, wound healing, and cell-cell signaling (Figure 6Aiii), suggesting a re-establishment of “social” micro-environmental interactions in the p50- and KPC1-overexpressing glioblastoma tumors (Bonavia et al., 2011). Downregulated processes included a reduced response to hypoxia required for maintaining glioblastoma stem cells (Heddleston et al., 2009), as well as reduced positive regulation of cell migration (Figure 6Aiii). Of the consistently changed genes, 21 are known NF- $\kappa$ B targets ( $p$  value  $< 3.4 \times 10^{-9}$ ; <http://www.bu.edu/nf-kb/gene-resources/target-genes/>). To further assess if the observed reduction in tumor size was the consequence of a reduction in proto-oncogenes and/or of an increased expression of tumor suppressor genes, we gathered gene annotations from various sources. Enrichment analysis on these gene annotations revealed a significant ( $p$  value  $< 1.4 \times 10^{-18}$ ) increase in the expression of 40 tumor suppressor genes, with no significant change in other classes (Figure 6B).

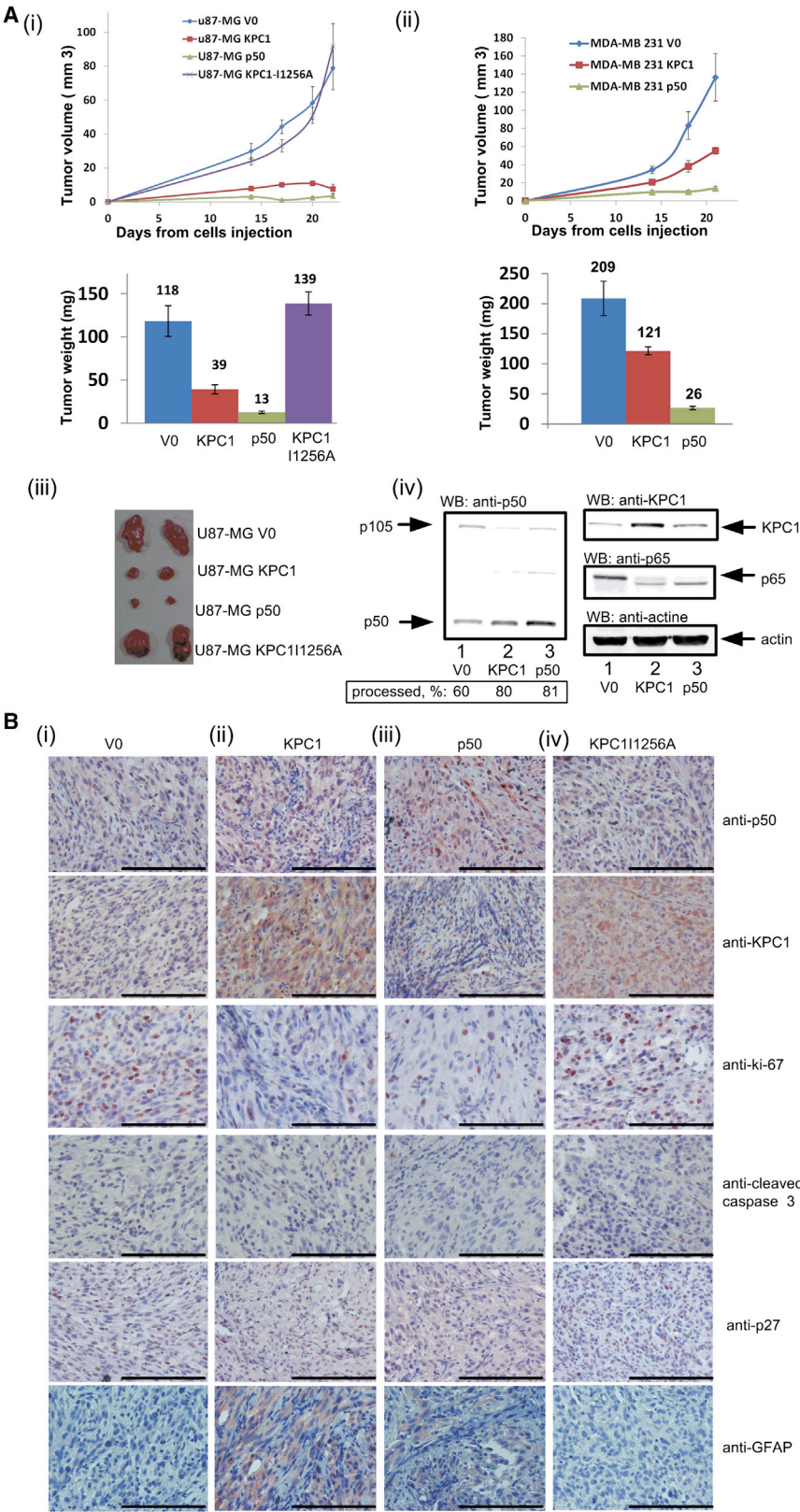
Finally, we integrated functional annotation enrichment and protein-protein interactions for the differentially regulated genes, which revealed a dense network of upregulated genes revolved around a few downregulated ones, such as interleukin-6 (IL-6), interleukin-6 receptor (IL-6R), and vascular endothelial growth factor A (VEGFA) (Figure 6C; Data S1). We included KPC1 and NF- $\kappa$ B in this analysis to retrieve possible known interactions, although KPC1 had no known interactions with any of the differentially regulated genes. Closer look at the core interaction network (Figure 6C, inset magnification) that included NF- $\kappa$ B is most prominently annotated with “regulation of cell migration” genes. Most other core network genes are upregulated and include many well-known tumor-suppressor genes.

Taken together, our findings strongly suggest a model of KPC1/p50 driven glioblastoma tumor growth inhibition, that centers around downregulated high mobility group protein HMGI-C (HMGA2), lin-28 homolog A (LIN28), IL-6/IL-6R, and VEGFA, and upregulated tumor suppressors, which in combination control the tumor-microenvironment as well as glioblastoma stem cell maintenance.

### Correlation between Expression of KPC1 and p50 in Human Tumoral and Normal Tissues

Finally, we examined the relationship between KPC1 and p50 in human tumors and normal tissues. Immunohistochemical staining of the two proteins (the two antibodies were shown to be specific; see Figure S6) revealed a high correlation between them in squamous cell carcinoma of head and neck (SCCHN, 52 sections;  $p$  value  $< 0.005$ ; Figure 7Aii), breast cancer (105 sections;  $p$  value  $< 0.0001$ ), and glioblastoma (192 sections;  $p$  value  $< 0.0017$ ) (Figure 7Ai). It should be emphasized though



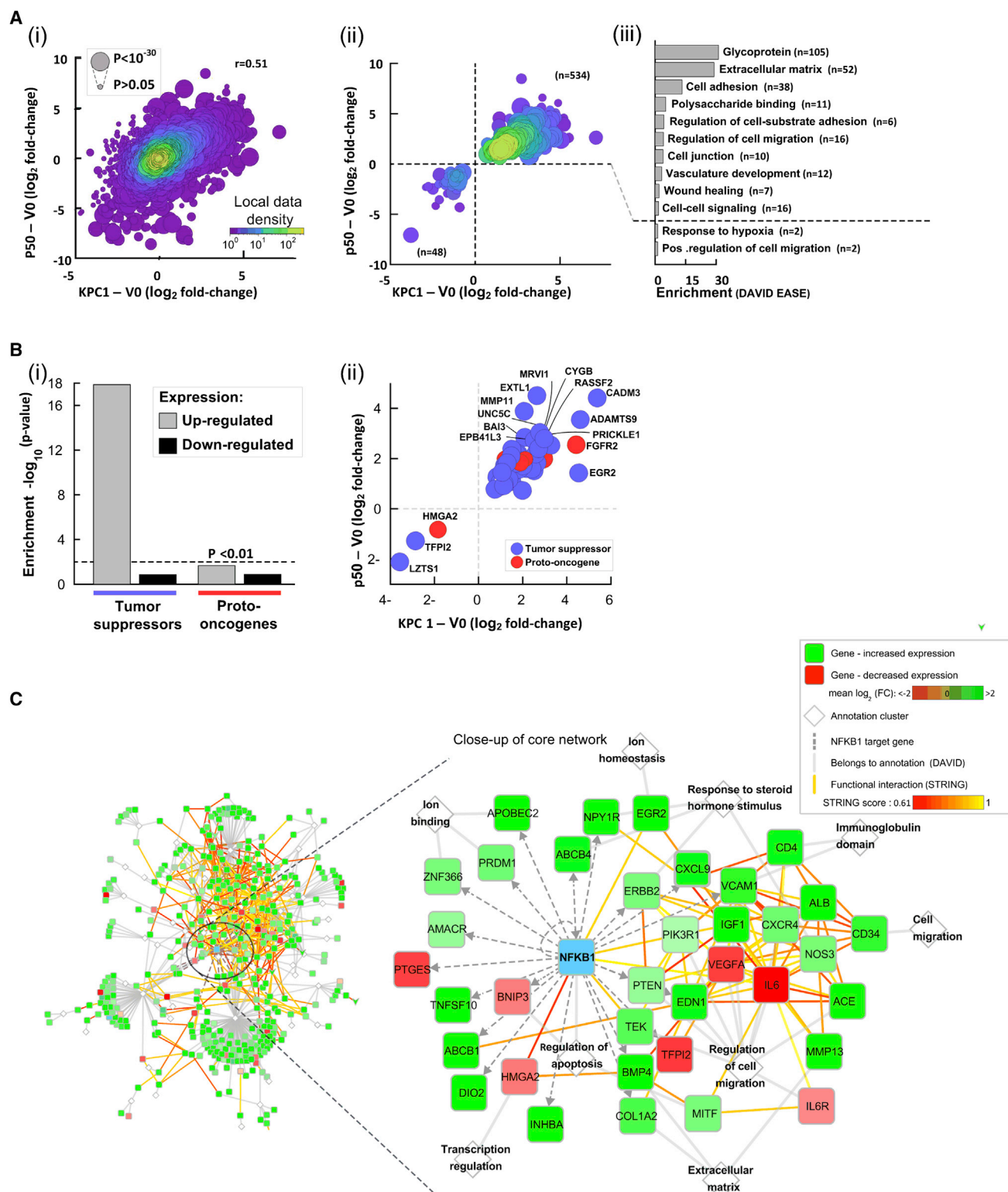


**Figure 5. KPC1-Mediated Excessive Generation of p50 Inhibits Tumor Growth**

(A) Growth rates and weights of tumor xenografts grown in mice, and derived from U87-MG (i) and MDA-MB 231 (ii) cells expressing V0, Myc-KPC1, and FLAG-p50. Data represent the mean of seven xenografts  $\pm$  SEM (iii) Tumors derived from U87-MG cells 3 weeks after inoculation. (iv) Enhanced generation of p50 and disappearance of p65 in tumors that overexpress KPC1 and p50. Proteins were resolved via SDS-PAGE, blotted onto nitrocellulose membrane, and detected using the appropriate antibodies. Processing was assessed as described under [Figure 2E](#). (B) Immunohistochemical staining of p50, KPC1, ki-67, cleaved caspase 3, p27Kip1, and GFAP in xenografts of U87-MG cells stably expressing V0 (i), Myc-KPC1 (ii), FLAG-p50 (iii), or KPC1I1256A (iv).

All scale bars, 100  $\mu$ m. Tumors were grown in mice and stained as described in the [Experimental Procedures](#). See also [Figure S5](#).





**Figure 6. RNA Analysis of Glioblastoma Xenografts Expressing KPC1 and p50**

(A) (i) Gene expression pattern as revealed by RNA-seq of U87-MG xenografts overexpressing either KPC1 or p50. Shown is log<sub>2</sub>-transformed fold-change in transcripts abundance against V0 control; dot size relates to the geometric mean significance over all experiments. (ii) Consistent downregulated and upregulated

(legend continued on next page)

that the linkage may be tumor-specific and not common to all patients with the “same” tumor.

To test the hypothesis that loss of KPC1 and nuclear p50 can be involved in the pathogenesis of malignant transformation, we analyzed the staining of the two proteins in SCCHN, breast cancer, and glioblastoma, and compared it to their staining in the normal parallel tissue. We observed a strong decrease in tumor samples stained for nuclear p50 compared to the healthy tissue (Figure 7B). As for KPC1, we observed a significant decrease in staining intensity (reflecting the amount of the protein) in cancerous compared to normal tissue in both SCCHN and glial cells, but not in breast cancer. Also, we noted a significant decrease in the number of tumor samples stained for KPC1 in SCCHN (Figure 7B). Taken together, these findings suggest that nuclear p50 is indeed a tumor suppressor lost in many malignancies. At least part of this p50 loss may be due to loss of KPC1 which catalyzes its formation, though this may not be common to all tumors.

## DISCUSSION

The vast majority of substrates of the Ub proteasome system are completely degraded. One intriguing and exceptional case is that of the p105 precursor of NF- $\kappa$ B that can be either completely degraded or processed in a limited manner to yield the p50 active subunit of the transcription factor. The “decision-making” mechanism resulting in one of the two distinct processes has remained largely elusive. The  $\beta$ TrCP Ub ligase has been identified as the tagging enzyme involved in complete degradation of p105, whereas the ligase involved in processing has remained unknown. We have now identified the KPC complex as the putative p105-processing ligase (Figures 1, 2, and 3).

Now that the two E3s involved in degradation and processing of p105 have been identified, it is still not known why ubiquitination by one enzyme results in a completely different fate of p105 than ubiquitination by the other and what determines the timing of the two reactions. It is possible that the two ligases catalyze the formation of chains that differ in their anchoring sites, length, and/or internal linkages. These, in turn, affect the recognition and mechanism of action of the 26S proteasome. As for timing, it can be that different physiological conditions and/or the degree of saturation of the ARs with bound p50s are involved in the “decision-making” process of whether the molecule will be processed or destroyed completely.

Studying the biological implications of manipulating KPC1 revealed that it suppresses anchorage-independent growth in a manner that is dependent on its ligase activity and the presence of p105. A corollary strong tumor-suppressive effect was demonstrated in xenografts of human tumors (Figures 4, 5, and

6). This effect is caused probably by a significant increase in an entire set of tumor suppressors, some of them like the brain-specific protein cell adhesion molecule 3 (CADM3) (Gao et al., 2009), was found inactivated in glioblastoma.

An important question relates to the transcriptional mechanism by which KPC1 and p50 exert their tumor-suppressive effect. An obvious assumption is that the stoichiometric excess of p50 generated by KPC1 would generate mostly p50-p50 homodimers rather than the “canonical” tumorigenic p50-p65 heterodimers. In line with this finding is also the observation that p65 level is decreased in KPC1-overexpressing as well as in p50-overexpressing xenografts (Figure 5Aiv). It appears that each dimer of NF- $\kappa$ B family has unique and even opposing biological function(s) and regulates a distinct subset of downstream genes (Siggers et al., 2012). p50 homodimer is supposed to act as a transcriptional repressor because it does not contain a transactivation domain (May and Ghosh, 1997). However, studies in vitro have shown that p50 homodimer can interact with different transcriptional modulators, such as Bcl-3 (Fujita et al., 1993), p300 (Deng and Wu, 2003), or HMGA1/2 (Perrella et al., 1999) that are involved in chromatin remodeling. Disproportionate p50 may shift the composition of NF- $\kappa$ B dimers, resulting in overall tumor-suppressive effect. Indeed, following overexpression of KPC1 or p50, there is a decrease in the level of what is probably the “canonical” tumorigenic NF- $\kappa$ B (p50-p65; Figure S5A).

Importantly, we found a strong correlation between the expression of KPC1 and that of p50 in human tumors (Figure 7A). Moreover, we found a significant decrease in nuclear p50 and KPC1 staining intensity in tumors compared to non-malignant tissue (Figure 7B). This observation suggests that loss of nuclear p50 may trigger malignant transformation. In line with these findings are data collected in the Catalog Of Somatic Mutations in Cancer (COSMIC) that show a significantly greater number of common tumors (e.g., breast, lung, CNS, and uterine cervix) with decreased expression of KPC1 transcripts compared to those with high expression (<http://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=RNF123&ln1=RNF123&start=1&end=1315&coords=AA%3AAA&sn=&ss=&hn=&sh=&samps=1001&expn=over&expn=under&id=4185>).

## EXPERIMENTAL PROCEDURES

### Materials, Plasmids, Expressed Proteins, and Cells

All materials (including plasmids and their construction, expression of proteins and their purification, and cultured cells and their manipulation), are described in the Extended Experimental Procedures.

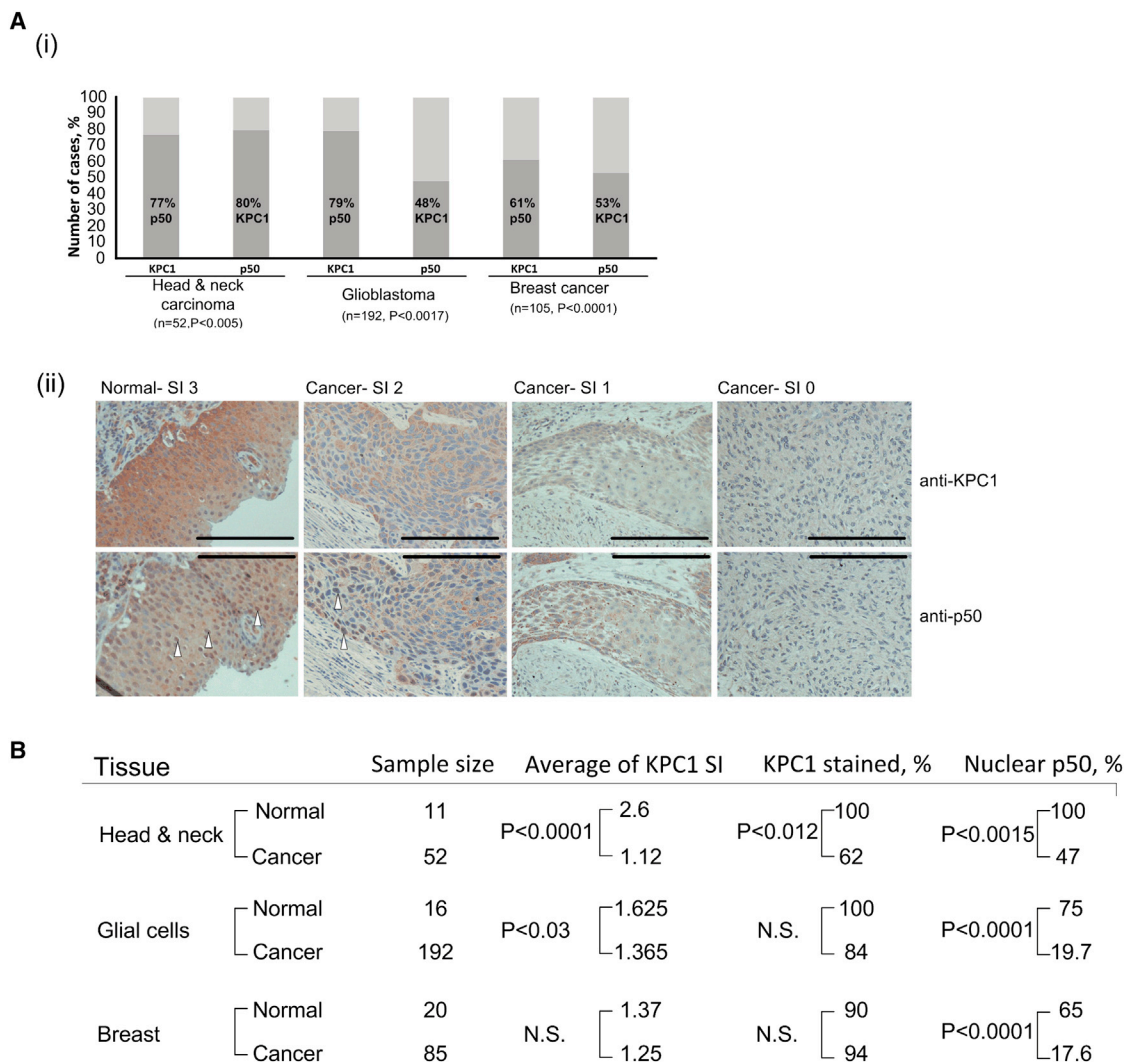
### Preparation and Fractionation of Crude Reticulocyte Lysate

Reticulocytes were induced in rabbits and lysates were prepared and fractionated over DEAE cellulose to Fraction I (unabsorbed material) and Fraction II

genes in the xenografts. Dot sizes are as in (i). (iii) Selected annotation clusters most enriched for either up- or downregulated genes (above or below dashed line, respectively).

(B) (i) Enrichment analysis of consistently upregulated and downregulated transcripts for tumor suppressors and proto oncogenes annotations. (ii) Expression differences for all tumor suppressors (blue) and proto oncogenes (red) from (i). Gene names of the strongest differentially regulated cancer-related genes are shown.

(C) Integrated analysis of functional annotation clusters and known functional and physical protein-protein interactions among all consistently upregulated and downregulated genes (green and red, respectively). NF- $\kappa$ B is shown in blue, and a close-up of the core interaction network surrounding NF- $\kappa$ B (inset) is displayed. See also Figure S5, Tables S1 and S2, and Data S1.



**Figure 7. Correlation between the Expression of KPC1 and p50 in Tumoral and Normal Human Tissues**

(A) (i) Correlation between expression of KPC1 and p50 in tumors. Immunohistochemistry of KPC1 and p50 in serial sections from SCCHN, and glioblastoma and breast cancer tissue arrays. P, p value. Analyses were carried out as described in the [Experimental Procedures](#). (ii) Representative immunostaining of SCCHN sections with anti-KPC1 or anti-p50. SI, staining intensity from 3 (strong positive) to 0 (negative). Arrowheads point to nuclear staining. Scale bars, 100  $\mu$ m.

(B) Statistical analysis of p50 and KPC1 staining in normal and cancerous head and neck, glial and breast tissues. "Average of KPC1 SI" represents mean of sample staining (number of samples is indicated under "Sample size"). "KPC1 stained, %" and "nuclear p50, %" represent percent of samples stained for KPC1 or nuclear p50. P, p value. The p value reflects the significance of difference between staining of normal and cancer tissue. SI, staining intensity; N.S., non-significant.

See also [Figure S6](#).

(high salt eluate) as described ([Hershko et al., 1983](#)). Fraction II (~200 mg) was further resolved using different successive chromatographic methods as described in the [Extended Experimental Procedures](#).

#### In Vitro Translation

p105 or p100 were translated in vitro in the presence of L-[ $^{35}$ S]methionine using the TNT T7 Quick reticulocyte lysate-based coupled transcription-translation kit according to the manufacturer's instructions.

#### In Vitro Conjugation and Processing of p105

Ub conjugation and processing of  $^{35}$ S-labeled p105 were carried out in a reconstituted cell-free system containing crude Fraction II as described

([Kravtsova-Ivantsiv et al., 2009](#)). For conjugation, 1  $\mu$ g of purified Kpc1-FLAG-TEV-6xHis, Kpc1I1256A-FLAG-TEV-6xHis, or 6xHis-KPC2 were added as indicated instead of Fraction II.

#### Ub Conjugates in Cells

HEK293 cells were transfected with control siRNA or siRNA against KPC1 as described above. After 24 hr, the cells were transfected with cDNAs coding for FLAG-p105 proteins along with cDNAs coding for HA-Ub and Myc-KPC1, or with an empty vector. After additional 24 hr, the proteasome inhibitor MG132 (20  $\mu$ M) was added for 3 hr, and the cells were lysed with RIPA buffer supplemented with freshly dissolved iodoacetamide and N-ethylmaleimide (5 mM each) to inhibit deubiquitinating enzymes. p105 (both free and

ubiquitinated) and free p50 were immunoprecipitated with immobilized anti-FLAG. The beads were washed five times with RIPA buffer and proteins were resolved by SDS-PAGE. Free and conjugated p105 (and free p50) were visualized using anti-FLAG.

### Tumorigenicity

Cell-based (soft agar) and animal (mice xenografts) tumorigenicity assays are described in the [Extended Experimental Procedures](#).

### RNA-Seq Analysis

RNA from U87-MG xenografts was isolated using RNA purification kit and analyzed using the Illumina HiSeq 2500 analyzer. Identification and clustering of the human genes are described in the [Extended Experimental Procedures](#).

### Immunohistochemistry and Statistical Analysis

The staining technique and statistical analysis of the staining data of SCCHN, breast cancer, and glioblastoma were performed as described in the [Extended Experimental Procedures](#).

### ACCESSION NUMBERS

The profile of gene expression in the tumors using RNA-seq analysis of transcripts reported in this paper is deposited in NCBI GEO under accession number GSE60530 and is available at <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60530>.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, two tables, and one data file and can be found with this article at <http://dx.doi.org/10.1016/j.cell.2015.03.001>

### AUTHOR CONTRIBUTIONS

A.C., T.S., and Y.T.K. contributed to the study design. Y.K.-I., I.S., and V.C.-K. contributed to the study design and carried out the biochemical and tumorigenicity assays. H.G. contributed to some of the biochemical experiments. T.Z. and A.A. performed the mass spectrometry analyses. I.N. and E.P. carried out the pathological observations. I.D. performed the statistical analyses of the stained human tumors. B.S. and G.S.-F. performed the computational analysis of the RNA-seq data. A.B. and M.J. synthesized the peptides and planned some of the experiments. A.C. supervised the study. Y.K.-I., I.S., and A.C. wrote the manuscript.

### ACKNOWLEDGMENTS

Research in the laboratory of A.C. is supported by grants from the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (AMRF), the Israel Science Foundation (ISF), the I-CORE Program of the Planning and Budgeting Committee and the ISF (Grant1775/12), the EU Treat PolyQ Network, the Nobel Laureates Invitation Program of Seoul National University, the Deutsch-Israelische Projektkooperation (DIP), and the Program for Targeting Cancer by Modulating Protein Dynamics supported by Albert Sweet (Malibu, CA). We thank Dr. Kazuhiro Iwai (Department of Molecular and Cellular Physiology, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto, Japan) for providing us with some cDNAs. RNA-seq and mapping to the human genome was done by the Technion Genome Center. B.S. is supported by a fellowship from the Swiss National Science Foundation. Y.T.K. is supported by The Basic Science Research Programs of the National Research Foundation of Korea (NRF-2013R1A2A2A01014170). A.C. is an Israel Cancer Research Fund (ICRF) USA Professor.

Received: August 15, 2014

Revised: December 29, 2014

Accepted: February 25, 2015

Published: April 9, 2015

### REFERENCES

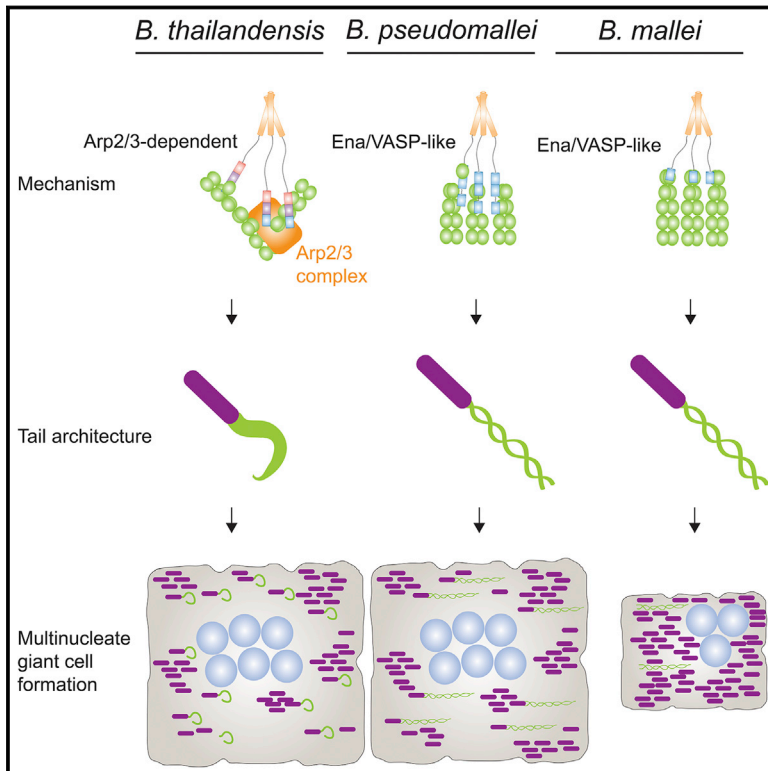
- Akiri, G., Cherian, M.M., Vijayakumar, S., Liu, G., Bafico, A., and Aaronson, S.A. (2009). Wnt pathway aberrations including autocrine Wnt activation occur at high frequency in human non-small-cell lung carcinoma. *Oncogene* 28, 2163–2172.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Barré, B., and Perkins, N.D. (2007). A cell cycle regulatory network controlling NF- $\kappa$ B subunit activity and function. *EMBO J.* 26, 4841–4855.
- Barré, B., Coqueret, O., and Perkins, N.D. (2010). Regulation of activity and function of the p52 NF- $\kappa$ B subunit following DNA damage. *Cell Cycle* 9, 4795–4804.
- Ben-Neriah, Y., and Karin, M. (2011). Inflammation meets cancer, with NF- $\kappa$ B as the matchmaker. *Nat. Immunol.* 12, 715–723.
- Betts, J.C., and Nabel, G.J. (1996). Differential regulation of NF- $\kappa$ B2(p100) processing and control by amino-terminal sequences. *Mol. Cell. Biol.* 16, 6363–6371.
- Bonavia, R., Inda, M.M., Cavenee, W.K., and Furnari, F.B. (2011). Heterogeneity maintenance in glioblastoma: a social network. *Cancer Res.* 71, 4055–4060.
- Carette, J.E., Raaben, M., Wong, A.C., Herbert, A.S., Obernosterer, G., Mulherkar, N., Kuehne, A.I., Kranzusch, P.J., Griffin, A.M., Ruthel, G., et al. (2011). Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* 477, 340–343.
- Cohen, S., Achtert-Weiner, H., and Ciechanover, A. (2004). Dual effects of I $\kappa$ B kinase  $\beta$ -mediated phosphorylation on p105 Fate: SCF<sup>( $\beta$ -TrCP)</sup>-dependent degradation and SCF<sup>( $\beta$ -TrCP)</sup>-independent processing. *Mol. Cell. Biol.* 24, 475–486.
- Cohen, S., Lahav-Baratz, S., and Ciechanover, A. (2006). Two distinct ubiquitin-dependent mechanisms are involved in NF- $\kappa$ B p105 proteolysis. *Biochem. Biophys. Res. Commun.* 345, 7–13.
- Deng, W.G., and Wu, K.K. (2003). Regulation of inducible nitric oxide synthase expression by p300 and p50 acetylation. *J. Immunol.* 171, 6581–6588.
- Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 4, 3.
- DiDonato, J.A., Mercurio, F., and Karin, M. (2012). NF- $\kappa$ B and the link between inflammation and cancer. *Immunol. Rev.* 246, 379–400.
- Fan, C.M., and Maniatis, T. (1991). Generation of p50 subunit of NF- $\kappa$ B by processing of p105 through an ATP-dependent pathway. *Nature* 354, 395–398.
- Fujita, T., Nolan, G.P., Liou, H.C., Scott, M.L., and Baltimore, D. (1993). The candidate proto-oncogene bcl-3 encodes a transcriptional coactivator that activates through NF- $\kappa$ B p50 homodimers. *Genes Dev.* 7, 1354–1363.
- Gao, J., Chen, T., Liu, J., Liu, W., Hu, G., Guo, X., Yin, B., Gong, Y., Zhao, J., Qiang, B., et al. (2009). Loss of NECL1, a novel tumor suppressor, can be restored in glioma by HDAC inhibitor-Trichostatin A through Sp1 binding site. *Glia* 57, 989–999.
- Hara, T., Kamura, T., Kotoshiba, S., Takahashi, H., Fujiwara, K., Onoyama, I., Shirakawa, M., Mizushima, N., and Nakayama, K.I. (2005). Role of the UBL-UBA protein KPC2 in degradation of p27 at G1 phase of the cell cycle. *Mol. Cell. Biol.* 25, 9292–9303.
- Heddleston, J.M., Li, Z., McLendon, R.E., Hjelmeland, A.B., and Rich, J.N. (2009). The hypoxic microenvironment maintains glioblastoma stem cells and promotes reprogramming towards a cancer stem cell phenotype. *Cell Cycle* 8, 3274–3284.
- Heissmeyer, V., Krappmann, D., Hatada, E.N., and Scheidereit, C. (2001). Shared pathways of I $\kappa$ B kinase-induced SCF<sup>( $\beta$ -TrCP)</sup>-mediated ubiquitination and degradation for the NF- $\kappa$ B precursor p105 and I $\kappa$ B $\alpha$ . *Mol. Cell. Biol.* 21, 1024–1035.
- Hershko, A., Heller, H., Elias, S., and Ciechanover, A. (1983). Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J. Biol. Chem.* 258, 8206–8214.



- Ivanov, V.N., Lee, R.K., Podack, E.R., and Malek, T.R. (1997). Regulation of Fas-dependent activation-induced T cell apoptosis by cAMP signaling: a potential role for transcription factor NF- $\kappa$ B. *Oncogene* 14, 2455–2464.
- Kamura, T., Hara, T., Matsumoto, M., Ishida, N., Okumura, F., Hatakeyama, S., Yoshida, M., Nakayama, K., and Nakayama, K.I. (2004). Cytoplasmic ubiquitin ligase KPC regulates proteolysis of p27<sup>Kip1</sup> at G1 phase. *Nat. Cell Biol.* 6, 1229–1235.
- Kravtsova-Ivantsiv, Y., Cohen, S., and Ciechanover, A. (2009). Modification by single ubiquitin moieties rather than polyubiquitination is sufficient for proteasomal processing of the p105 NF- $\kappa$ B precursor. *Mol. Cell* 33, 496–504.
- Lin, L., and Ghosh, S. (1996). A glycine-rich region in NF- $\kappa$ B p105 functions as a processing signal for the generation of the p50 subunit. *Mol. Cell. Biol.* 16, 2248–2254.
- Lin, L., DeMartino, G.N., and Greene, W.C. (1998). Cotranslational biogenesis of NF- $\kappa$ B p50 by the 26S proteasome. *Cell* 92, 819–828.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- MacKichan, M.L., Logeat, F., and Israël, A. (1996). Phosphorylation of p105 PEST sequence via a redox-insensitive pathway up-regulates processing of p50 NF- $\kappa$ B. *J. Biol. Chem.* 271, 6084–6091.
- May, M.J., and Ghosh, S. (1997). Rel/NF- $\kappa$ B and I $\kappa$ B proteins: an overview. *Semin. Cancer Biol.* 8, 63–73.
- Mercer, J., Snijder, B., Sacher, R., Burkard, C., Bleck, C.K., Stahlberg, H., Pelkmans, L., and Helenius, A. (2012). RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. *Cell Rep.* 2, 1036–1047.
- Mercurio, F., Zhu, H., Murray, B.W., Shevchenko, A., Bennett, B.L., Li, J., Young, D.B., Barbosa, M., Mann, M., Manning, A., and Rao, A. (1997). IKK-1 and IKK-2: cytokine-activated I $\kappa$ B kinases essential for NF- $\kappa$ B activation. *Science* 278, 860–866.
- Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* 108, 193–199.
- Orian, A., Gonen, H., Bercovich, B., Fajerman, I., Eytan, E., Israël, A., Mercurio, F., Iwai, K., Schwartz, A.L., and Ciechanover, A. (2000). SCF<sup>( $\beta$ -TrCP)</sup> ubiquitin ligase-mediated processing of NF- $\kappa$ B p105 requires phosphorylation of its C-terminus by I $\kappa$ B kinase. *EMBO J.* 19, 2580–2591.
- Palombella, V.J., Rando, O.J., Goldberg, A.L., and Maniatis, T. (1994). The ubiquitin-proteasome pathway is required for processing the NF- $\kappa$ B1 precursor protein and the activation of NF- $\kappa$ B. *Cell* 78, 773–785.
- Perkins, N.D. (2012). The diverse and complex roles of NF- $\kappa$ B subunits in cancer. *Nat. Rev. Cancer* 12, 121–132.
- Perrella, M.A., Pellacani, A., Wiesel, P., Chin, M.T., Foster, L.C., Ibanez, M., Hsieh, C.M., Reeves, R., Yet, S.F., and Lee, M.E. (1999). High mobility group-I(Y) protein facilitates NF- $\kappa$ B binding and transactivation of the inducible nitric-oxide synthase promoter/enhancer. *J. Biol. Chem.* 274, 9045–9052.
- Pikarsky, E., and Ben-Neriah, Y. (2006). NF- $\kappa$ B inhibition: a double-edged sword in cancer? *Eur. J. Cancer* 42, 779–784.
- Rahman, M.M., and McFadden, G. (2011). Modulation of NF- $\kappa$ B signalling by microbial pathogens. *Nat. Rev. Microbiol.* 9, 291–306.
- Salmerón, A., Janzen, J., Soneji, Y., Bump, N., Kamens, J., Allen, H., and Ley, S.C. (2001). Direct phosphorylation of NF- $\kappa$ B1 p105 by the I $\kappa$ B kinase complex on serine 927 is essential for signal-induced p105 proteolysis. *J. Biol. Chem.* 276, 22215–22222.
- Senftleben, U., Cao, Y., Xiao, G., Greten, F.R., Krahn, G., Bonizzi, G., Chen, Y., Hu, Y., Fong, A., Sun, S.C., and Karin, M. (2001). Activation by IKK $\alpha$  of a second, evolutionary conserved, NF- $\kappa$ B signaling pathway. *Science* 293, 1495–1499.
- Siggers, T., Chang, A.B., Teixeira, A., Wong, D., Williams, K.J., Ahmed, B., Ragoussis, J., Udalo, I.A., Smale, S.T., and Bulky, M.L. (2012). Principles of dimer-specific gene regulation revealed by a comprehensive characterization of NF- $\kappa$ B family DNA binding. *Nat. Immunol.* 13, 95–102.
- Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39, D561–D568.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- UniProt Consortium (2013). Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acid Res.* 41, D43–D47.
- Voce, D.J., Schmitt, A.M., Uppal, A., McNerney, M.E., Bernal, G.M., Cahill, K.E., Wahlstrom, J.S., Nassiri, A., Yu, X., Crawley, C.D., et al. (2014). NF- $\kappa$ B1 is a haploinsufficient DNA damage-specific tumor suppressor. *Oncogene*. <http://dx.doi.org/10.1038/onc.2014.211>.
- Zaaroor-Regev, D., de Bie, P., Scheffner, M., Noy, T., Shemer, R., Heled, M., Stein, I., Pikarsky, E., and Ciechanover, A. (2010). Regulation of the polycomb protein Ring1B by self-ubiquitination or by E6-AP may have implications to the pathogenesis of Angelman syndrome. *Proc. Natl. Acad. Sci. USA* 107, 6788–6793.
- Zhao, M., Sun, J., and Zhao, Z. (2013). TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res.* 41, D970–D976.

# Virulent *Burkholderia* Species Mimic Host Actin Polymerases to Drive Actin-Based Motility

## Graphical Abstract



## Authors

Erin L. Benanti, Catherine M. Nguyen, Matthew D. Welch

## Correspondence

ebenanti@aduro.com (E.L.B.),  
welch@berkeley.edu (M.D.W.)

## In Brief

Pathogenic and non-pathogenic *Burkholderia* species use distinct strategies to drive actin-based motility in host cells, suggesting that mimicry of different polymerization mechanisms influences bacterial spread from cell to cell and the outcome of infection.

## Highlights

- Orthologs of BimA mimic host Arp2/3 activators or Ena/VASP actin polymerases
- BimA orthologs promote distinct actin organization and motility parameters
- The efficiency of initiating motility is key to cell fusion and spread
- Pathogen mimicry of a spectrum of actin assembly mechanisms influences infection



# Virulent *Burkholderia* Species Mimic Host Actin Polymerases to Drive Actin-Based Motility

Erin L. Benanti,<sup>1,2,\*</sup> Catherine M. Nguyen,<sup>1</sup> and Matthew D. Welch<sup>1,\*</sup>

<sup>1</sup>Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

<sup>2</sup>Present address: Aduro Biotech, 626 Bancroft Way 3C, Berkeley, CA 94710, USA

\*Correspondence: [ebenanti@aduro.com](mailto:ebenanti@aduro.com) (E.L.B.), [welch@berkeley.edu](mailto:welch@berkeley.edu) (M.D.W.)

<http://dx.doi.org/10.1016/j.cell.2015.02.044>

## SUMMARY

*Burkholderia pseudomallei* and *B. mallei* are bacterial pathogens that cause melioidosis and glanders, whereas their close relative *B. thailandensis* is non-pathogenic. All use the trimeric autotransporter BimA to facilitate actin-based motility, host cell fusion, and dissemination. Here, we show that BimA orthologs mimic different host actin-polymerizing proteins. *B. thailandensis* BimA activates the host Arp2/3 complex. In contrast, *B. pseudomallei* and *B. mallei* BimA mimic host Ena/VASP actin polymerases in their ability to nucleate, elongate, and bundle filaments by associating with barbed ends, as well as in their use of WH2 motifs and oligomerization for activity. Mechanistic differences among BimA orthologs resulted in distinct actin filament organization and motility parameters, which affected the efficiency of cell fusion during infection. Our results identify bacterial Ena/VASP mimics and reveal that pathogens imitate the full spectrum of host actin-polymerizing pathways, suggesting that mimicry of different polymerization mechanisms influences key parameters of infection.

## INTRODUCTION

The pseudomallei group of *Burkholderia* species are Gram-negative bacteria that include the opportunistic human pathogens *B. pseudomallei* (*Bp*), which causes the severe disease melioidosis, and *B. mallei* (*Bm*), a clonal descendant of *Bp* that primarily infects equine but can cause acute human disease (Cheng and Currie, 2005; Wilkinson, 1981). *Bp* and *Bm* are of heightened concern because they are resistant to numerous antibiotics, spread via an aerosol route, and exhibit low infectious doses. A third species, *B. thailandensis* (*Bt*), is closely related yet is not pathogenic to humans. All three share key virulence factors despite their differences in infectivity. Because *Bt* infection in animals and cells recapitulates many features of *Bp* and *Bm* virulence, it has been used as a model system to study *Burkholderia* pathogenesis (Galyov et al., 2010; Haraga et al., 2008; West et al., 2008). However, the basis for the dramatic differences in *Burkholderia* virulence remains a mystery. Under-

standing these differences will provide insight into the evolution of virulence among closely related bacteria.

Key features of *Burkholderia* virulence include their ability to invade host cells and escape from phagosomes into the cytosol, where they replicate and undergo actin-based motility (Kespi-chayawattana et al., 2000; Stevens et al., 2005a). As with other pathogens, *Burkholderia* use actin-based motility to facilitate cell-to-cell spread, which enables dissemination within hosts while evading the immune system (Goldberg, 2001). However, *Burkholderia* spread is distinct in that it can occur through bacterial-mediated fusion of host cells (French et al., 2011; Kespi-chayawattana et al., 2000) induced by type VI secretion (Schwarz et al., 2014; Toesca et al., 2014), rather than the engulfment of membrane protrusions. Regardless of the spread mechanism, motility drives bacteria to the host plasma membrane to facilitate spread.

To enable actin-based motility, pathogens express proteins that mimic or activate host actin-nucleating factors to promote the polymerization of actin monomers (G-actin) into filaments (F-actin) that elongate at fast-growing barbed ends. *Listeria monocytogenes* and *Shigella flexneri* produce mimics or activators of host nucleation-promoting factors (NPFs) to stimulate the host Arp2/3 complex (Welch and Way, 2013), which constructs branched actin filament networks (Campellone and Welch, 2010). *Rickettsia* species use an NPF that activates the Arp2/3 complex early in infection (Reed et al., 2014) and later use a mimic of host formins (Haglund et al., 2010; Kleba et al., 2010; Madasu et al., 2013), which nucleate filaments and processively bind to barbed ends to enhance elongation rates (Paul and Pollard, 2009). *Vibrio* and *Chlamydia* species, which do not undergo motility, express proteins that use tandem actin monomer-binding sequences to nucleate F-actin (Jewett et al., 2006; Liverman et al., 2007; Namgoong et al., 2011; Pernier et al., 2013; Tam et al., 2007; Yu et al., 2011). Despite this diversity, no pathogens have been shown to mimic host Ena/VASP proteins, which are weak nucleators, bind barbed ends, enhance filament elongation, and bundle filaments (Barzik et al., 2005; Breitsprecher et al., 2011; Hansen and Mullins, 2010; Krause et al., 2003; Samarin et al., 2003; Winkelman et al., 2014). Within this context, it is unclear whether *Burkholderia* mimics or activates host nucleators or nucleates actin by an unknown mechanism. Additionally, the consequences of different nucleation mechanisms on actin-based motility and infection have not been examined.

*Burkholderia* actin-based motility requires the *Burkholderia* intracellular motility A (BimA) protein (Schell et al., 2007; Sitthidet

et al., 2010; Stevens et al., 2005b), a member of the trimeric autotransporter (AT) family. Trimeric ATs contain highly conserved C-terminal sequences that mediate secretion, localization to the outer membrane, and trimerization (Cotter et al., 2005; Dautin and Bernstein, 2007). These include a  $\beta$  barrel, as well as an adjacent  $\sim 35$  residue  $\alpha$  helix (Stevens et al., 2005b), which in other trimeric ATs forms a trimeric coiled coil that is positioned inside the barrel (Meng et al., 2006). Trimeric ATs also contain N-terminal passenger domains that are exposed on the bacterial surface. Sequence comparisons suggest that the passenger domains of BimA from *Bt*, *Bp*, and *Bm* polymerize host actin by different mechanisms (Stevens et al., 2005b). BtBimA contains putative actin-binding WASP homology 2 (WH2) and Arp2/3-binding central and acidic (CA) motifs that are conserved among NPFs that activate the Arp2/3 complex (Campellone and Welch, 2010). Previous work showed that the BtBimA CA motifs are required to activate Arp2/3 nucleation in vitro and promote actin association in host cells (Sitthidet et al., 2010). In contrast, BpBimA and BmBimA contain three (Bp) or one (Bm) putative WH2 motifs and lack CA sequences. Although two of the three putative BpBimA WH2 sequences were shown to be required for actin binding, nucleation, and plaque formation (Sitthidet et al., 2011), the activity of BmBimA has not been examined. It also remains unclear what molecular mechanisms BpBimA and BmBimA use to nucleate actin and the role trimerization plays in this process.

Here, we investigated the mechanisms of actin assembly and motility driven by BimA orthologs from different *Burkholderia* species. Despite the conservation of key virulence factors (French et al., 2011; Galyov et al., 2010; Haraga et al., 2008), we found that *Bt*, *Bp*, and *Bm* have evolved divergent mechanisms for actin polymerization. These differences result in *Burkholderia* that generate actin tails with distinct filament architectures and exhibit different efficiencies of motility and host-cell fusion, providing one potential explanation for the increased virulence of *Bp* and *Bm* relative to *Bt*. These observations demonstrate that intracellular pathogens employ the full spectrum of actin assembly pathways and suggest that distinct mechanisms allow microbes to fine-tune motility to control spread during infection.

## RESULTS

### BtBimA Activates the Host Arp2/3 Complex to Nucleate Branched Filaments, whereas BpBimA and BmBimA Independently Nucleate Unbranched Filaments

Previous studies indicated that BtBimA requires its CA motifs and the host Arp2/3 complex for actin nucleation, whereas BpBimA nucleates actin independent of Arp2/3 (Sitthidet et al., 2010; Stevens et al., 2005b). The activity of BmBimA was not investigated. In these studies, incomplete BtBimA and BpBimA passenger domains that lacked trimeric coiled coils were fused to GST, which would induce non-native dimerization (Sitthidet et al., 2010, 2011; Stevens et al., 2005b). The activity of these constructs was modest ( $\sim 1.5$ - to  $2.5$ -fold increased polymerization rates). We hypothesized that the full passenger domains in their native oligomeric states would exhibit higher nucleation activities that would be dependent (BtBimA) or independent

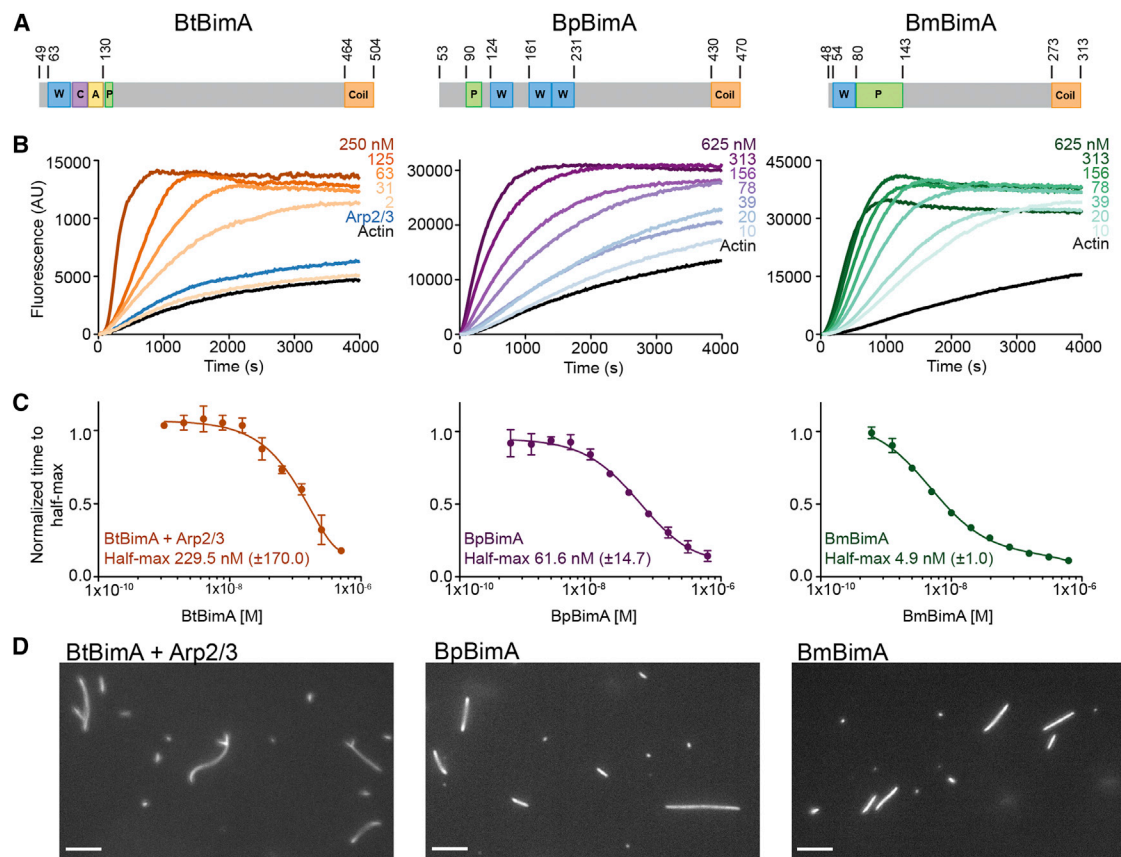
(BpBimA, BmBimA) of Arp2/3 complex. To test this, we purified versions of each ortholog that spanned the passenger domain and extended through the trimeric coiled coil (Figure 1A) (Meng et al., 2006; Szczesny and Lupas, 2008). We used pyrene actin-polymerization assays to test each purified BimA for activity. BtBimA alone had little nucleation activity (Figures S1A and S1B), yet it activated the Arp2/3 complex in a concentration-dependent manner ( $>5$ -fold increase relative to actin alone), indicating that it functions as an NPF (Figures 1B and 1C). In contrast, BpBimA and BmBimA displayed robust nucleation activity ( $>7$ - or  $8$ -fold increase relative to actin alone), and Arp2/3 addition had no effect (Figure S1B). To compare their potencies, we determined the time it took to reach half-maximum fluorescence over a range of BimA concentrations and fit the resulting curves (Figure 1C). All three orthologs are potent nucleators with 230 nM BtBimA (with 50 nM Arp2/3), 60 nM BpBimA, and 5 nM BmBimA required for half-maximum activity. Consistent with its ability to activate the Arp2/3 complex, BtBimA and Arp2/3 generated branched actin filaments (Figure 1D). In contrast, BpBimA and BmBimA generated unbranched filaments. Thus, BtBimA is an NPF that activates Arp2/3, whereas BpBimA and BmBimA independently nucleate actin, similar to host proteins in the formin, tandem monomer nucleator, and Ena/VASP families.

### BpBimA and BmBimA Bind Filament Barbed Ends, Processively Elongate Filaments, and Remove Capping Protein from Filaments

Formins and Ena/VASP proteins processively track barbed ends and alter their elongation rates (Paul and Pollard, 2009), whereas tandem monomer nucleators have not been shown to track growing filament ends (Carlier et al., 2011; Namgoong et al., 2011; Pernier et al., 2013). To compare BimA behavior to that of members of these families, we investigated how BimA orthologs interact with filament ends and tested whether they affect elongation. We labeled BpBimA and BmBimA with Alexa Fluor 488 at Cys residues that were introduced at their C termini and used total internal reflection fluorescence (TIRF) microscopy to monitor the effect of BimA on rhodamine-labeled actin polymerization. BmBimA and BpBimA bound and remained associated with barbed ends as they grew (Figure 2A; Movies S1 and S2). Growth rate measurements of barbed ends in the presence or absence of BimA indicated that BmBimA and BpBimA increased elongation rates by  $3.5$ - to  $4.5$ -fold relative to actin alone. To measure their apparent affinities for barbed ends, we monitored elongation of pre-polymerized actin seeds with increasing BimA concentrations and fit the concentration dependence of the decrease in time to half-maximum fluorescence (Figure 2B). These measurements produced apparent affinities for barbed ends of 350 nM for BpBimA and 3 nM for BmBimA, similar to the barbed-end affinities of formin and Ena/VASP proteins (Hansen and Mullins, 2010; Harris et al., 2004; Otomo et al., 2005; Pruyne et al., 2002; Winkelman et al., 2014).

Ena/VASP proteins are further distinguished from formins by the ability of at least one family member (Ena) to bundle and elongate two filaments at once (Winkelman et al., 2014). BmBimA and BpBimA also frequently gathered two or even three filaments and mediated their simultaneous elongation (Figures 2C





**Figure 1. BtBimA Mimics Host NPFs, whereas BpBimA and BmBimA Independently Nucleate Actin**

(A) Domain schematics of BimA from different *Burkholderia* species. Actin-related and oligomerization sequences are shown. Numbers refer to amino acid position in full-length proteins. W, WASP homology 2; C, central; A, acidic; P, proline rich; Coil, trimeric coiled coil.

(B) Pyrene actin polymerization reactions with increasing BimA concentrations with Arp2/3 complex (Bt) or in its absence (Bp, Bm).

(C) The time to half-maximum fluorescence of the polymerization curves in (B) normalized to actin alone. The means  $\pm$  SD are shown, and fits of each data set are reported to estimate the concentration at which half-maximum activity is observed.

(D) Epifluorescence images of polymerization reactions containing BtBimA with Arp2/3 complex, BpBimA, or BmBimA after 15 min that were stabilized and stained with rhodamine phalloidin.

Scale bars, 5  $\mu$ m. See also Figure S1.

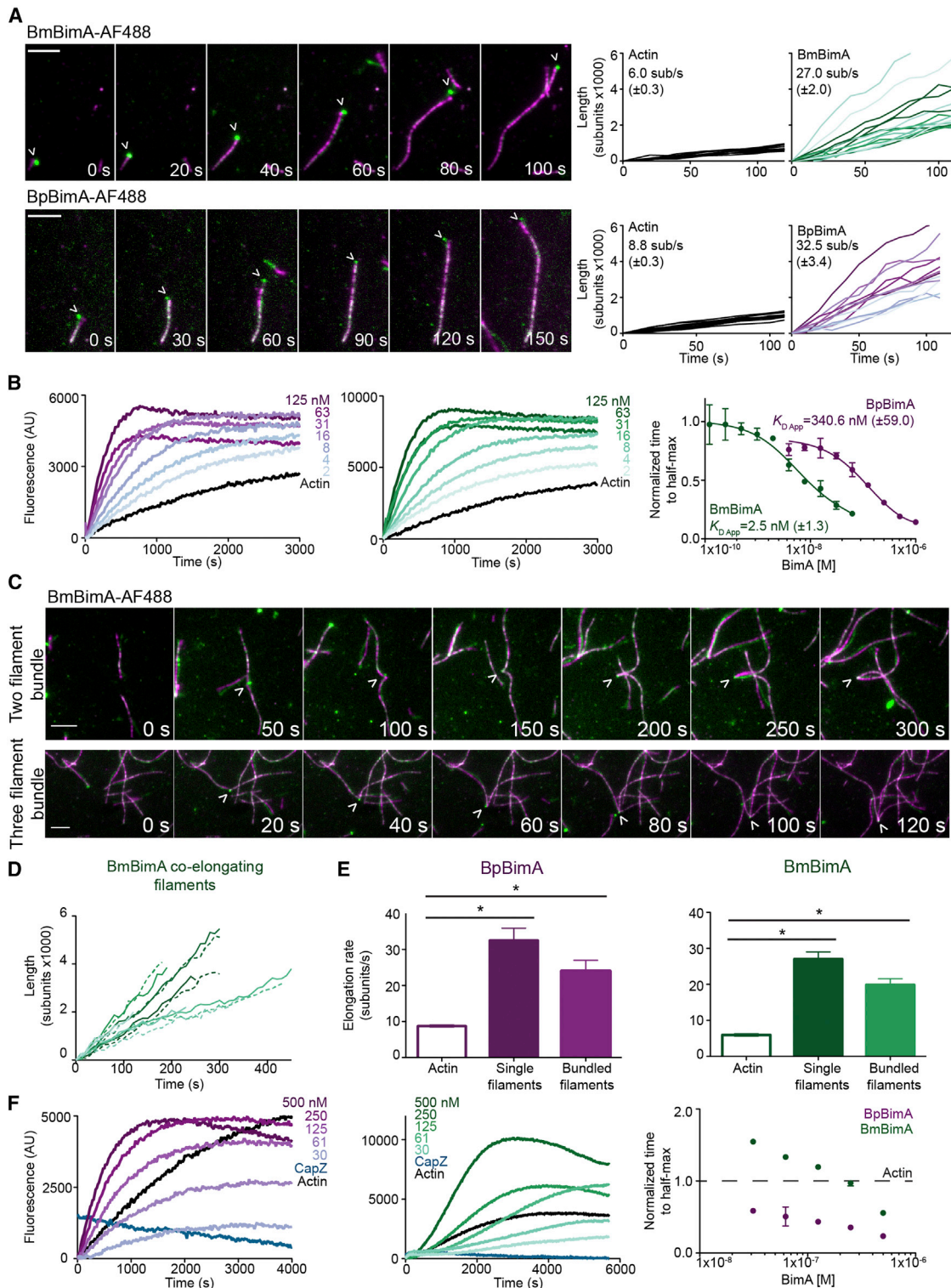
and S2; Movies S3, S4, S5, and S6). Bundled filaments bound by a single BimA spot grew at similar rates relative to each other (Figure 2D) and to individual BimA-elongated filaments (Figure 2E). Thus, in this regard, BpBimA and BmBimA activity closely resembles that of Ena/VASP proteins.

In cells, barbed ends are often capped by CapZ, which prevents filament elongation (Cooper and Sept, 2008). Purified CapZ displays a high affinity of  $\sim 2$  nM for barbed ends (Caldwell et al., 1989) and a slow dissociation rate from barbed ends ( $\sim 30$  min half-life) (Schafer et al., 1996). Formins and Ena/VASP proteins compete with CapZ to prevent capping (Barzik et al., 2005; Breitsprecher et al., 2008; Paul and Pollard, 2009) but have not been reported to remove CapZ from pre-capped filaments. To test whether BpBimA and BmBimA have anti-capping activity, we monitored their ability to promote filament elongation when added to pre-formed actin seeds simultaneously with CapZ (data not shown) or when added 5 min after filaments had been pre-capped with CapZ (Figure 2F). Under

both conditions, 10 nM CapZ alone completely inhibited elongation, whereas the addition of 30 nM BpBimA or BmBimA to pre-capped filaments allowed filament elongation. These results suggest that BpBimA and BmBimA have the unusual ability to remove CapZ from barbed ends to enable elongation. Together, our results demonstrate that BpBimA and BmBimA are barbed-end-binding proteins that increase elongation rates, bundle filaments, and have anti-capping activity, similar to host Ena/VASP actin polymerases.

### Oligomerization Is Required for BpBimA and BmBimA Activity

One feature of Ena/VASP proteins required for their actin polymerase activity is a tetrameric coiled coil located at the C terminus (Breitsprecher et al., 2008; Hansen and Mullins, 2010; Kühnel et al., 2004). Based on the functional similarities of BpBimA and BmBimA with Ena/VASP proteins, we predicted that trimerization is important for BimA barbed-end-binding and elongation

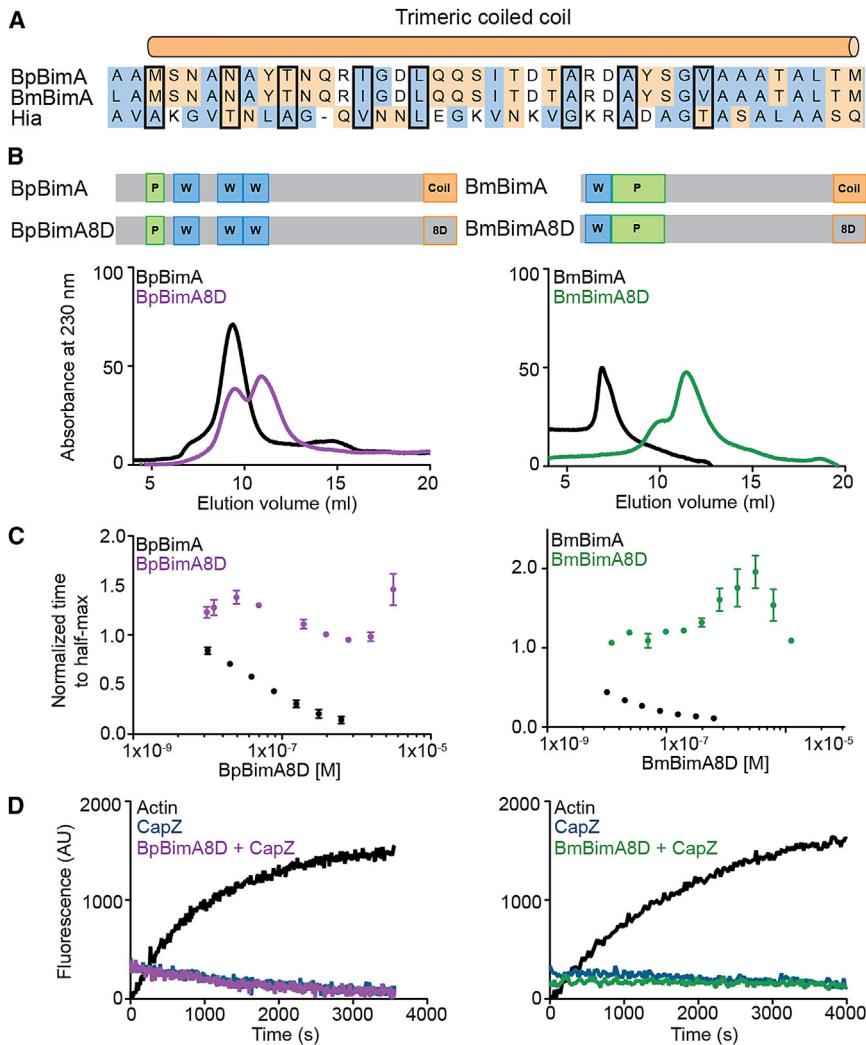


**Figure 2. BmBimA and BpBimA Processively Bind Growing Filament Barbed Ends, Increase Elongation Rates, Bundle Filaments, and Outcompete CapZ for Barbed-End Binding**

(A) Left, TIRF images showing BmBimA-AF488 and BpBimA-AF488 (green) and rhodamine-labeled actin (magenta). Time (s) is indicated. Scale bars, 3  $\mu$ m. Right, filament length (number of subunits  $\times$  1,000) over time for a minimum of 10 filaments with the mean elongation rate (sub/s  $\pm$  SEM) listed.

(B) Left and middle, pyrene elongation assays with a range of BimA concentrations. Right, the time to half-maximum fluorescence normalized to actin alone for BmBimA (green) or BpBimA (purple) with the mean apparent  $K_D \pm$  SD listed.

(legend continued on next page)



**Figure 3. BpBimA and BmBimA Oligomerization Is Required for Actin Nucleation and Barbed-End Binding**

(A) Alignment of the trimeric coiled coils from BpBimA and BmBimA with the Hia trimeric coiled coil (Meng et al., 2006). Positions replaced with Asp residues in the BimA8D mutants are outlined in black. Hydrophobic residues are blue, and charged residues orange.

(B) Domain schematics of wild-type and BimA8D mutants in which eight positions were changed to Asp residues are shown above gel-filtration elution profiles of wild-type and mutant proteins.

(C) The time to half-maximum fluorescence normalized to actin alone in polymerization reactions with increasing BimA8D proteins. The means  $\pm$  SD are shown with the wild-type data from Figure 1C for reference.

(D) Elongation reactions with 10 nM CapZ and with or without BimA8D proteins.

relative to their wild-type counterparts by gel filtration (Figure 3B), suggesting that the higher oligomeric state of the wild-type proteins was disrupted and that the trimeric coiled coil mediates oligomerization.

To determine the importance of oligomerization in actin nucleation and elongation, we compared the activity of each BimA8D mutant with wild-type BimA in pyrene actin assembly assays. Both BpBimA8D and BmBimA8D lacked detectable nucleation activity, and at high concentrations, both mutants inhibited polymerization, suggesting that they sequester G-actin (Figure 3C). Neither mutant could relieve CapZ inhibition

of elongation at concentrations two orders of magnitude higher than that at which wild-type BimA relieved inhibition (Figure 3D). Thus, oligomerization of BpBimA and BmBimA is required for actin nucleation, barbed-end elongation, and anti-capping activity, similar to the requirement of tetramerization for Ena/VASP actin polymerase activity.

### WH2 Motifs Are Required for BpBimA and BmBimA Activity

Ena/VASP barbed-end binding and filament elongation require two WH2 sequences (also called the globular actin-binding or GAB and filamentous actin-binding or FAB motifs) (Bachmann

activities. To estimate the oligomeric state of wild-type and mutant BimA proteins, we determined the behavior of wild-type BimA on a gel-filtration column. BpBimA and BmBimA eluted at volumes corresponding to globular proteins of much larger molecular weight (>500 kDa) than expected based on their sequences (127 kDa for BpBimA; 88 kDa for BmBimA; Figure 3B). However, other trimeric ATs display aberrant behavior that is attributed to their elongated structures (Cotter et al., 2005; Hartmann et al., 2012; Mack et al., 1994). Next, we replaced eight residues of the predicted  $\alpha$  helix with aspartic acid (D) to disrupt the formation of the trimeric coiled coil (Figure 3A). The resulting BpBimA8D and BmBimA8D mutants eluted at smaller sizes

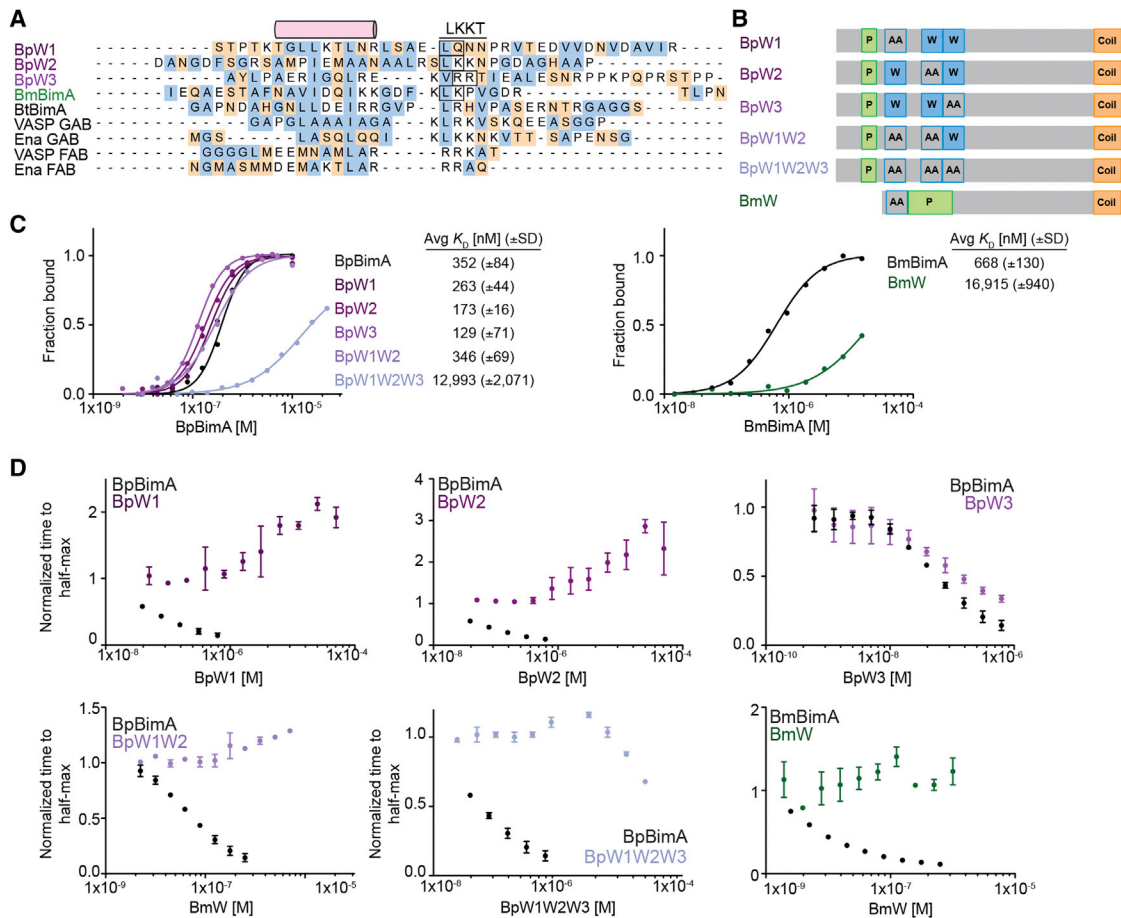
(C) TIRF images of BmBimA-AF488 (arrowhead) elongating two (top) or three (bottom) filaments (colors as in A).

(D) Graph of filament length (number of subs  $\times$  1,000) over time for pairs of co-elongating filaments (each pair is in matched colors of solid and dashed lines). (E) Mean elongation rates (sub/s  $\pm$  SEM) for actin alone or for single filaments or two- and three-filament bundles elongated with BmBimA-AF488 (left, green) or BpBimA-AF488 (right, purple). Asterisks denote paired samples that are significantly different (one-way ANOVA;  $p < 0.0001$ ).

(F) Left and middle, elongation reactions first incubated with 10 nM CapZ, followed by BimA and G-actin addition. Right, the times to half-maximum fluorescence normalized to actin alone (dashed black line) for BmBimA (green) or BpBimA (purple) are shown with the means  $\pm$  SD.

See also Figure S2 and Movies S1, S2, S3, S4, S5, and S6.





**Figure 4. WH2 Requirements for BpBimA and BmBimA G-Actin Binding and Nucleation**

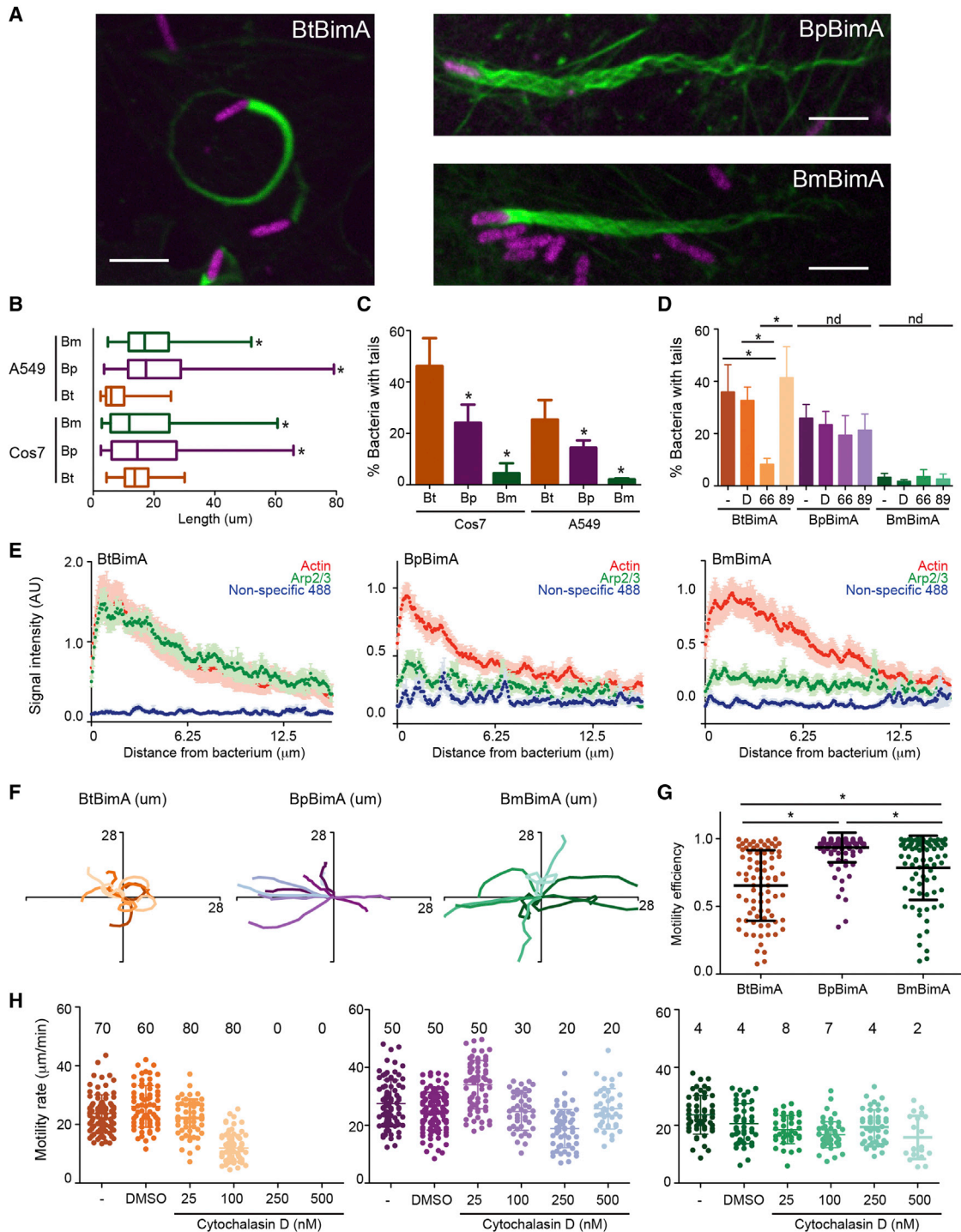
(A) Alignment of predicted WH2 sequences from BimA orthologs with known WH2 (GAB and FAB) sequences from human VASP and *Drosophila* Ena. The conserved  $\alpha$  helix and LKKT motif are indicated. Hydrophobic residues are blue, and charged residues orange. Boxed residues were mutated to AA. (B) Domain schematics of WH2 mutant BimA proteins. Predicted WH2 motifs that were mutated are outlined in blue and denoted by AA. (C) Anisotropy measurements of monomeric actin488 binding to BimA. Data are represented by circles, and Hill equation fits are shown as solid lines. The means  $K_D \pm SD$  from at least two experiments are listed. (D) The time to half-maximum fluorescence normalized to actin alone in polymerization reactions with BimA. The means  $\pm SD$  are shown with wild-type data from Figure 1C for reference.

et al., 1999; Hansen and Mullins, 2010). BpBimA and BmBimA contain three (Bp) or one (Bm) putative WH2 sequences (Figure 1A) (Stevens et al., 2005a, 2005b) that are implicated in actin binding, and we hypothesized that each BimA would use these sequences for nucleation and elongation, similar to Ena/VASP proteins. We determined whether the BimA WH2 sequences are competent for actin binding by replacing two positions of the conserved LKKT signature sequence with alanines (Figures 4A and 4B). To measure G-actin binding, we titrated increasing wild-type or WH2 mutant BimA into a fixed concentration of Alexa Fluor 488-labeled G-actin (actin488) in low ionic strength buffer to prevent actin polymerization (Figure 4C). Anisotropy measurements with increasing BimA produced saturable binding curves that were fit using the Hill equation. Wild-type BpBimA and BmBimA bound actin488 with apparent affinities of 350 nM and 670 nM (Figure 4C). Mutating any single predicted WH2 sequence in BpBimA (BpW1, BpW2, and BpW3 mutants) or

mutating the first two WH2 motifs together (BpW1W2) had little effect on G-actin binding (Figure 4C). However, simultaneously mutating all three WH2 sequences in the BpW1W2W3 mutant severely reduced binding ( $K_D \sim 13 \mu M$ ). The single BmW mutant also exhibited a significant reduction in affinity relative to wild-type BmBimA ( $K_D \sim 17 \mu M$ ). These data demonstrate that BmBimA contains one WH2 motif capable of binding G-actin and suggest that BpBimA contains up to three WH2 motifs that bind G-actin.

To determine whether BimA WH2 motifs are required for filament nucleation and elongation, we used the pyrene actin assembly assay to test each WH2 mutant for its ability to nucleate actin. Although the BpW1 and BpW2 mutants bound G-actin with high affinity, they lacked nucleation activity, as evidenced by the absence of a decrease in time to half-maximum fluorescence intensity (Figure 4D). Notably, these mutants inhibited polymerization at concentrations above  $\sim 1 \mu M$ , which is likely





**Figure 5. BimA from Different *Burkholderia* Species Mediates the Formation of Distinct Actin Tails and Parameters of Actin-Based Motility**  
 (A) Merged images showing Cos7 cells infected with different *Bt* strains that constitutively express RFP (magenta). F-actin was stained with Alexa Fluor 488 phalloidin (green). Scale bars, 5  $\mu\text{m}$ .  
 (B) Mean actin-tail lengths for the indicated strains in Cos7 and A549 cells. Boxes outline the 25<sup>th</sup> and 75<sup>th</sup> percentiles, midlines denote the medians, and whiskers show minimum and maximum lengths. Asterisks denote lengths significantly different from BtBimA in that cell type (one-way ANOVA;  $p < 0.05$ ).  
 (C) Percent bacteria with actin tails ( $\pm$ SD).  
 (D) Percent bacteria with actin tails ( $\pm$ SD) without or with 1 hr treatment with DMSO, the Arp2/3 inhibitor CK-666, or the inactive compound CK-689. nd, no difference.

(legend continued on next page)

due to remaining intact WH2 motifs sequestering G-actin. In contrast, the activity of BpW3 was similar to that of wild-type BpBimA, indicating that the third WH2 motif is not required for activity. BpW1W2 and BpW1W2W3 were unable to nucleate actin, and BpW1W2W3 also lacked inhibitory activity, consistent with the low affinity of this mutant for G-actin (Figure 4C). Thus, although all three BpBimA WH2 motifs may bind actin monomers, only the N-terminal two WH2 motifs are required for nucleation activity. For BmBimA, the sole WH2 mutant BmW lacked detectable nucleation activity and did not inhibit actin polymerization, indicating that this WH2 motif is required for nucleation (Figure 4D). Together with the requirement for oligomerization, these results suggest that BpBimA uses up to six and BmBimA uses up to three WH2 motifs within a BimA trimer to mediate actin nucleation and elongation. Thus, like Ena/VASP proteins, BpBimA and BmBimA require oligomerization and WH2 motifs for their actin polymerase activities.

### Mechanistic Differences among BimA Orthologs Expressed in *Bt* Result in Distinct Filament Organization and Parameters of Actin-Based Motility

To compare how differences in the biochemical properties of BimA orthologs affect actin filament organization and actin-based motility by *Burkholderia*, we replaced the endogenous copy of *Bt bimA* with an identical copy of *Bt bimA* or with *bimA* from *Bp* or *Bm*. To assess BimA synthesis and localization, we engineered strains producing internally FLAG-tagged versions of BimA. Bacteria producing FLAG-BimA were compared with those expressing the corresponding untagged BimA orthologs in plaque assays and formed identical numbers and sizes of plaques as the untagged versions (data not shown). We found that FLAG-BtBimA and FLAG-BmBimA were produced at similar levels, whereas FLAG-BpBimA was produced at 2- to 2.5-fold higher levels than the others (Figure S3A). All displayed polar localization similar to endogenous BimA in *Bp* (Figure S3B) (Sitthidet et al., 2011; Stevens et al., 2005b) and also enabled actin-tail formation by *Bt* following infection of tissue culture cells (Figure 5A), similar to previous results for BimA orthologs in *Bp* (Stevens et al., 2005a). Unless otherwise noted, data were obtained using strains expressing untagged BimA.

Confocal microscopy of actin-tail structures generated by each strain indicated that tails produced by BtBimA were curved and consisted of a dense actin network (Figure 5A; FLAG-BimA images are shown). In contrast, BpBimA and BmBimA produced longer, straighter tails that consisted of bundled F-actin strands (Figure 5B). Interestingly, the proportion of BpBimA bacteria associated with tails was moderately reduced and that of BmBimA bacteria was severely reduced relative to BtBimA bacteria (Figure 5C), suggesting differences in motility initiation. The Arp2/3 complex localized specifically to BtBimA tails, but not BpBimA or BmBimA tails, as monitored by immunofluorescence

staining and by the presence of Arp3-GFP (Figures 5E and S3D–S3F). Furthermore, treatment of infected cells with the Arp2/3 complex inhibitor CK-666 (but not the control compound CK-689) reduced the frequency of tail formation by BtBimA, whereas tails formed by BpBimA or BmBimA were unaffected (Figure 5D). Thus, BtBimA requires Arp2/3 complex activity to produce shorter and curved actin tails, whereas the Ena/VASP mimics BpBimA and BmBimA work independent of Arp2/3 to produce longer tails with bundled filaments.

We next used time-lapse imaging to compare actin-based motility parameters among strains producing different BimA orthologs. Once movement was initiated, all three strains exhibited similar average velocities of  $\sim 30 \mu\text{m}/\text{min}$ , with rates varying widely for each strain (Figure S3C). However, the movement paths of the BtBimA strain were more curved, whereas those of BpBimA and BmBimA bacteria were straighter (Figure 5F; Movie S7). To quantify path straightness and motility efficiency, we divided the linear displacement by the total distance traveled over 40 s for each bacterium. By this criteria, the BpBimA and BmBimA strains moved more efficiently than the BtBimA strain (Figure 5G). We also monitored motility following treatment with increasing concentrations of cytochalasin D (CD), a drug that binds to and inhibits growth of barbed ends. Although treatment with 250 or 500 nM CD reduced the frequency of motility for all strains, it completely blocked movement by the BtBimA strain, whereas some BpBimA and BmBimA bacteria resisted treatment and moved at rates similar to those in untreated cells (Figure 5H). This is consistent with the ability of these proteins to relieve CapZ inhibition. Overall, these results demonstrate that Arp2/3-dependent and Ena/VASP-like mechanisms of actin nucleation and elongation can drive similar rates of actin-based motility. However, distinct motility mechanisms also result in differences in motility initiation, paths, efficiency, and susceptibility to inhibition by actin-disrupting drugs.

### The Mechanism of BimA Actin Assembly Impacts the Efficiency of Host Cell Fusion

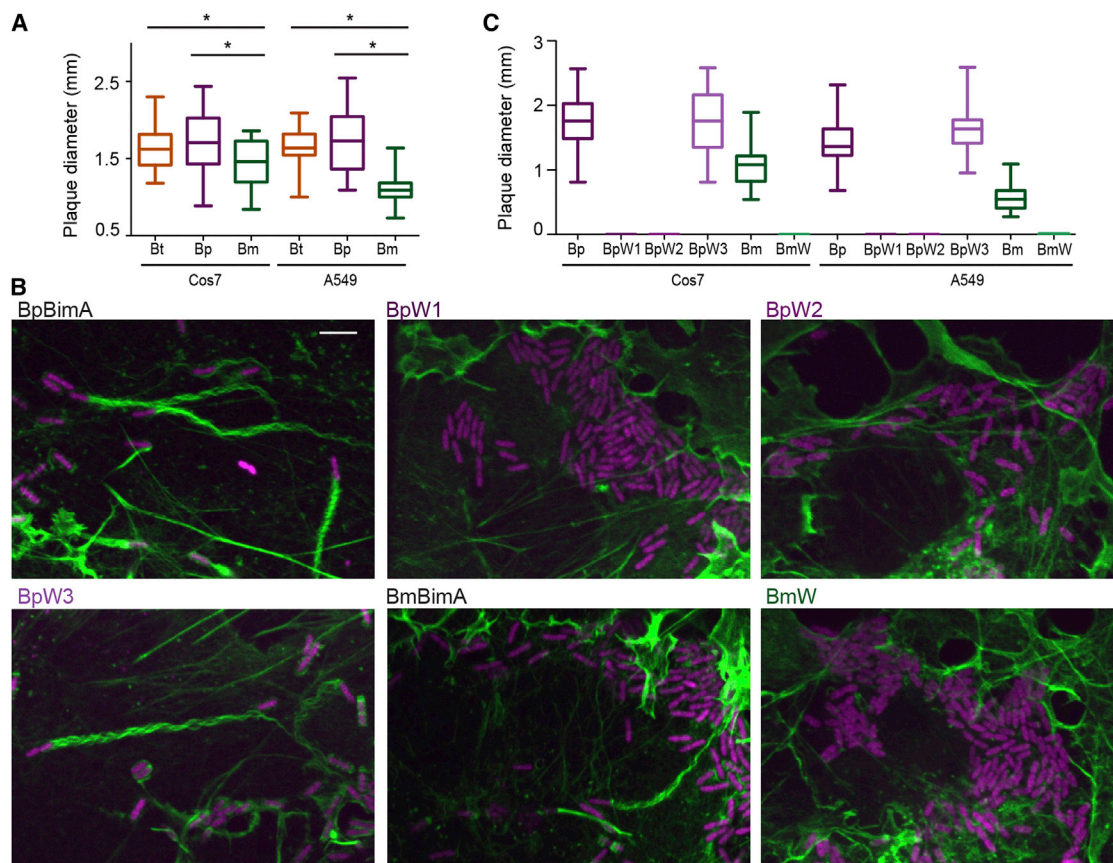
Bacterial-mediated host cell fusion is essential for *Burkholderia* virulence, and fusion depends on actin-based motility (French et al., 2011; Schwarz et al., 2014). We therefore hypothesized that differences in actin polymerization mechanisms and motility impact *Burkholderia*-mediated fusion. To compare fusion efficiencies, strains producing different BimA orthologs were assessed for their ability to form plaques on host cell monolayers. BtBimA and BpBimA strains formed similarly sized plaques, but the BmBimA strain produced smaller plaques in both Cos7 and A549 cells (Figure 6A). This plaque defect did not correlate with a reduction in BmBimA expression, a defect in localization (Figures S3A and S3B), intracellular replication (data not shown), slower rate of motility (Figure 5D), or differences in movement paths (Figure 5F). Instead, the reduced plaque size strongly correlated

(E) Mean actin, Arp2/3, or non-specific 488 nm fluorescence intensities ( $\times 1,000$ ) from at least 10 actin tails ( $\pm$ SEM) are plotted along the first 12.5  $\mu\text{m}$  of tail.

(F) Tracks (ten per strain) depicting motility over 100 s for bacteria in Cos7 cells.

(G) Motility efficiency from at least 80 tracks per strain, calculated as described in the text.

(H) Motility velocity over 40 s for each strain in Cos7-Lifeact-EGFP cells untreated or treated with DMSO or CD. Mean frequencies of motile bacteria are listed. For all panels except (B), asterisks denote paired samples that are statistically different from one another (one-way ANOVA;  $p < 0.0001$ ). See also Figure S3 and Movie S7.



**Figure 6. BimA Mechanisms of Actin Nucleation Impact Actin Tail Formation and Host Cell Fusion**

(A) Plaque diameters in Cos7 or A549 cell monolayers infected with *Bt* expressing wild-type BpBimA, BpBimA, or BmBimA.

(B) Merged images of Cos7 cells infected with wild-type or mutant BimA-expressing *Bt* strains that constitutively express RFP (magenta). F-actin was stained with Alexa Fluor 488 phalloidin (green). Scale bars, 5  $\mu$ m.

(C) Plaque diameters in Cos7 or A549 cell monolayers infected with *Bt* expressing wild-type or mutant BimA.

(A and C) Boxes outline the 25<sup>th</sup> and 75<sup>th</sup> percentiles, midlines denote the medians, and whiskers show minimum and maximum tail lengths. Asterisks denote significantly different sample pairs (one-way ANOVA;  $p < 0.05$ ). See also Figure S4.

with a low frequency of actin-tail association (<5%; Figure 5C), suggesting that the initiation of actin-based motility is a crucial parameter for *Burkholderia* host cell fusion.

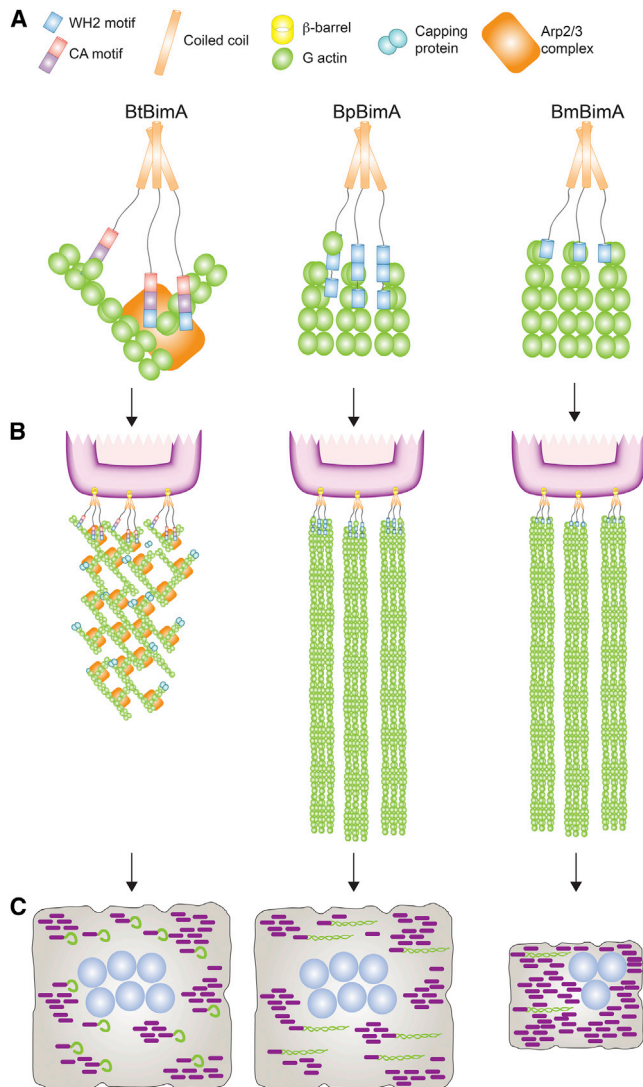
We investigated the mechanisms used by BpBimA and BmBimA to enable intracellular motility and fusion by testing the prediction that each BimA would require one or more WH2 motifs to drive movement during infection. We generated *Bt* strains in which we replaced endogenous *bimA* with genes encoding WH2 mutants (including BpW1, BpW2, BpW3, and BmW). The synthesis and localization of BimA in each strain were similar to those of wild-type BimA as determined by immunofluorescence staining against internal FLAG tags (Figure S4). Consistent with their lack of polymerization activity, the BpW1 and BpW2 mutants had no detectable association with F-actin, whereas the BpW3 mutant generated actin tails similar to wild-type BpBimA (Figure 6B). The BmW mutant similarly lacked actin association. We next examined the efficiency of fusion by measuring the plaque-forming ability of the mutants. The BpW1, BpW2, and BmW strains were completely defective in

plaque formation, whereas the BpW3 strain generated a similar number and size of plaques as the wild-type BpBimA strain (Figure 6C). Thus, BpBimA requires WH2 motifs W1 and W2, and BmBimA requires its sole WH2 motif for actin-based motility and fusion during *Bt* infection. These results demonstrate that molecular mimicry of host Ena/VASP proteins is crucial for actin-based motility driven by BpBimA and BmBimA in host cells.

## DISCUSSION

Microbe-driven actin-based motility plays a crucial role in spread and virulence of many pathogens and is also used as a model system to understand actin dynamics in host cells. Almost all motile pathogens use an actin-polymerization mechanism that depends on the host Arp2/3 complex (Truong et al., 2014; Welch and Way, 2013), with the exception of *Rickettsia* species that also use a formin-like mechanism late in infection (Haglund et al., 2010; Madasu et al., 2013; Reed et al., 2014). Here we





**Figure 7. A Model for BtBimA, BpBimA, and BmBimA Nucleation and Formation of Actin Tails and Multinucleated Giant Cells**

(A) BtBimA activates host Arp2/3 complex using its WCA domain to nucleate branched actin networks. BpBimA nucleates and elongates filaments from barbed ends using its two N-terminal WH2 motifs, whereas BmBimA exhibits the same activities using its single WH2 motif to generate bundled filaments. (B) BtBimA and Arp2/3 form shorter tails with branched actin networks. BpBimA and BmBimA produce longer tails of bundled filaments. (C) *Burkholderia* synthesizing BtBimA or BpBimA form more actin tails and generate larger multinucleated giant cells compared with bacteria producing BmBimA.

show that virulent *Burkholderia* species employ a previously undescribed mechanism to co-opt actin for motility, mimicking Ena/VASP actin polymerases to nucleate, elongate, and bundle filaments.

In contrast to BtBimA activation of the Arp2/3 complex, we found that BpBimA and BmBimA function independently of Arp2/3. Both directly nucleate actin, bind processively to filament barbed ends, and increase the elongation rate of bundled

filaments (Figure 7). Moreover, they have the unusual ability to dissociate CapZ from barbed ends, a property they share with the *Vibrio cholerae* VopF protein (Pernier et al., 2013), as well as to resist the action of the drug CD during infection. These activities most closely mirror those of the Ena/VASP family of actin polymerases (Krause et al., 2003), which similarly bind to filament barbed ends, processively elongate filaments, and under some conditions protect growing barbed ends from CapZ inhibition (Barzik et al., 2005; Bear et al., 2002; Breitsprecher et al., 2008; Hansen and Mullins, 2010; Schirenbeck et al., 2006; Winkelmann et al., 2014). Perhaps the most striking parallel between the activities of these proteins is their shared ability to gather and elongate multiple filaments simultaneously, which is a recently described property of *Drosophila* Ena (Winkelmann et al., 2014). This enables BpBimA and BmBimA to generate networks of bundled filaments similar to those formed by Ena/VASP proteins in mammalian cells (Krause et al., 2003).

BpBimA and BmBimA also display sequence requirements similar to those of Ena/VASP for their activity, including both WH2 and coiled coils. BimA trimerization would result in the close positioning of multiple WH2 sequences (three for BmBimA and up to nine for BpBimA), which may promote elongation through the proximity of G-actin and barbed-end-binding sites (Figure 7). Higher-order oligomerization of BpBimA or BmBimA may also contribute to their ability to bundle and elongate multiple filaments, similar to the tetramerization requirement and role of clustering in Ena/VASP protein processivity (Bachmann et al., 1999; Breitsprecher et al., 2008; Hansen and Mullins, 2010; Samarin et al., 2003).

Our analysis also revealed that *Bt* producing different BimA orthologs generate actin tails with distinct organization (Figure 7) and move differently within host cells. BtBimA bacteria generate shorter and curvier tails that are densely packed with actin. Although they move at velocities similar to bacteria expressing other orthologs, they exhibited lower motility efficiency due to their curvier paths. These characteristics are similar to those of *L. monocytogenes*, which use the NPF ActA to activate Arp2/3 and polymerize actin (Welch et al., 1998). In contrast, *Bt* synthesizing BpBimA or BmBimA produce longer, straighter tails consisting of bundled filaments and exhibit a higher motility efficiency due to straighter paths. The bundled filament organization in these tails is consistent with the Ena/VASP-like activities of BpBimA and BmBimA in vitro. Thus, virulent *Burkholderia* species have evolved a distinct molecular mechanism for efficient actin-based motility.

Actin-based motility driven by BimA orthologs also differs in promoting host cell fusion. In particular, the BmBimA strain exhibited reduced fusion efficiency relative to the BtBimA and BpBimA strains. Reduced fusion did not correlate with altered velocity, directionality of movement, or even the underlying polymerization mechanism. Fusion did, however, correlate with a ~5- to 10-fold reduction in the frequency of BmBimA-driven tail formation and motility initiation. This defect is surprising considering that BmBimA possesses more potent nucleation and elongation activity than BpBimA in vitro. Thus, despite its increased activity in isolation, BmBimA appears to be a less efficient actin nucleator in cells. Differences between the activity of BimA outside versus inside cells may be due to differential



requirements for other bacterial or host factors that play important roles in motility. For example, BmBimA may require additional host actin regulators that are present in equine cells but absent from human cells. Regardless, these findings suggest that the ability of BimA to initiate actin-based motility is a particularly crucial parameter to enable host cell fusion.

Why has BimA evolved distinct mechanisms of generating actin filaments? *Bp* is a soil saprophyte that infects a broad range of mammals (Galyov et al., 2010), and *Bm* is a clonal descendent of *Bp* that has undergone reductive genome evolution and cannot survive outside of a host (Galyov et al., 2010; Godoy et al., 2003; Nierman et al., 2004). *Bt* is a soil saprophyte like *Bp*, but its natural host is unknown. The sequence and mechanistic similarities between BpBimA and BmBimA reflect their evolutionary lineage, and it is interesting that the length and complexity of BmBimA have decreased relative to those of BpBimA, mirroring changes in the sizes of their respective genomes. The evolution of BtBimA is less clear. We speculate that distinct BimA mechanisms evolved in response to different host cell environments; whereas BpBimA is optimized for infection of various mammals, BmBimA is fine-tuned to function in equines, and BtBimA has adapted to unknown hosts in the soil environment.

As a whole, our results highlight that even closely related species have evolved to mimic a diverse spectrum of host actin-polymerizing pathways and that mimicry of different polymerization mechanisms may influence key parameters of infection, such as host cell fusion and bacterial dissemination. The acquisition of distinct polymerization mechanisms may be relevant for the evolution of virulence, as both highly virulent *Burkholderia* species evolved mimics of host Ena/VASP proteins that nucleate, elongate, and bundle actin filaments, as well as remove filament capping proteins. However, the ramifications of these actin-regulatory capabilities on virulence remain to be appreciated. Studying the divergent actin-based motility mechanisms of closely related species represents a powerful approach to unravel the evolution of pathogenic strategies for exploiting actin and to reveal new principles that govern the generation, dynamics, and regulation of distinct actin networks in cells.

## EXPERIMENTAL PROCEDURES

### Protein Purification and Fluorescent Labeling

SUMO-6XHis-BimA orthologs were expressed in *E. coli* and isolated by affinity, anion-exchange, and/or gel-filtration chromatography. Alexa Fluor 488 C<sub>5</sub> maleimide (Life Technologies) was used to label Cys-containing BpBimA and BmBimA. Details of the purification and labeling methods are in the [Extended Experimental Procedures](#).

### Bulk Actin Assembly Assays

Pyrene actin polymerization reactions contained rabbit skeletal muscle G-actin (1 or 3  $\mu$ M, 10% pyrene labeled) and BimA. Elongation assays were performed by mixing F-actin with BimA and initiated by adding G-actin (250 nM). For CP competition, 10 nM CP was mixed with F-actin seeds, BimA was added after 5 min, and reactions were initialized with G-actin (250 nM). For buffers and further details, see [Extended Experimental Procedures](#).

### BimA Gel-Filtration Chromatography

BimA proteins were run over a Superdex 200 10/300 GL gel filtration column (GE Healthcare) and detected by absorbance at 230 nm. See [Extended Experimental Procedures](#) for further details.

### Epifluorescence and TIRF Microscopy to Visualize Actin Filaments

For epifluorescence microscopy, polymerization reactions were stabilized by adding 1  $\mu$ M rhodamine-phalloidin (Life Technologies), and F-actin was visualized as described previously (Haglund et al., 2010). TIRF microscopy was performed as previously described (Kovar et al., 2006), except that reactions contained 1  $\mu$ M ATP-actin (33% rhodamine-labeled). At least 10 filaments without BimA, or 20 filaments with 100 pM BpBimA or 10 nM BmBimA, were measured over 240 s by manually tracing filaments. See [Extended Experimental Procedures](#) for further details.

### G-Actin Binding Anisotropy

BimA proteins were added to 100 nM Alexa Fluor 488-labeled G-actin under non-polymerizing conditions. Binding curves were fit using the Hill equation, and the means of at least two titrations are reported. See [Extended Experimental Procedures](#) for further details.

### Bacterial Strain Construction

Parental *Bt* strain E264 was a gift from P. Cotter (University of North Carolina, Chapel Hill, NC, USA). *Bt bimA* and *Bt motA2* strains were gifts from C. French and J. F. Miller (University of California, Los Angeles, CA, USA). Details of strain construction are in the [Extended Experimental Procedures](#) and in [Tables S1](#) and [S2](#).

### *B. thailandensis* Infection of Mammalian Cells

A549, COS7, HEK293T, and U2OS cells were from the University of California, Berkeley tissue culture facility. Polyclonal cell lines stably expressing Lifeact-EGFP (COS7) or F-Tractin-Wasabi (A549) were generated by lentiviral transduction. For infections, bacteria grown in LB broth were added to pre-seeded mammalian cells for 1 hr at 37°C at high MOIs (100–200) before adding media with gentamicin (Gm; 0.5 mg/ml). Fixed and live-cell analyses were performed with samples at 8–15 hpi. For plaque assays, cells were infected (MOI of 0.1) for 1 hr and overlaid with a mixture of agarose in media with Gm and imaged by neutral red staining at 36 hpi. See [Extended Experimental Procedures](#) for further details.

### Image Analysis

Image analysis details are in the [Extended Experimental Procedures](#).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, two tables, and seven movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.044>.

## AUTHOR CONTRIBUTIONS

E.L.B. and M.D.W. designed the experiments; E.L.B. and C.M.N. generated the data; E.L.B. analyzed the data; and E.L.B. and M.D.W. wrote the paper.

## ACKNOWLEDGMENTS

We thank Peggy Cotter, Jeffrey Iwig, and Rebecca Lamason for comments on the manuscript and Tiago Barros for expertise in model design. We are grateful to Melissa Anderson, Erin Garcia, Peggy Cotter, Todd French, and Jeff F. Miller for reagents and technical advice and to Taro Ohkawa for imaging assistance. E.L.B. was supported by NIH/NIGMS NRSA F32 GM093652. M.D.W. is supported by NIH/NIGMS R01 GM059609 and NIH/NIAID R01 AI109044.

Received: October 24, 2014

Revised: December 22, 2014

Accepted: February 5, 2015

Published: April 9, 2015

## REFERENCES

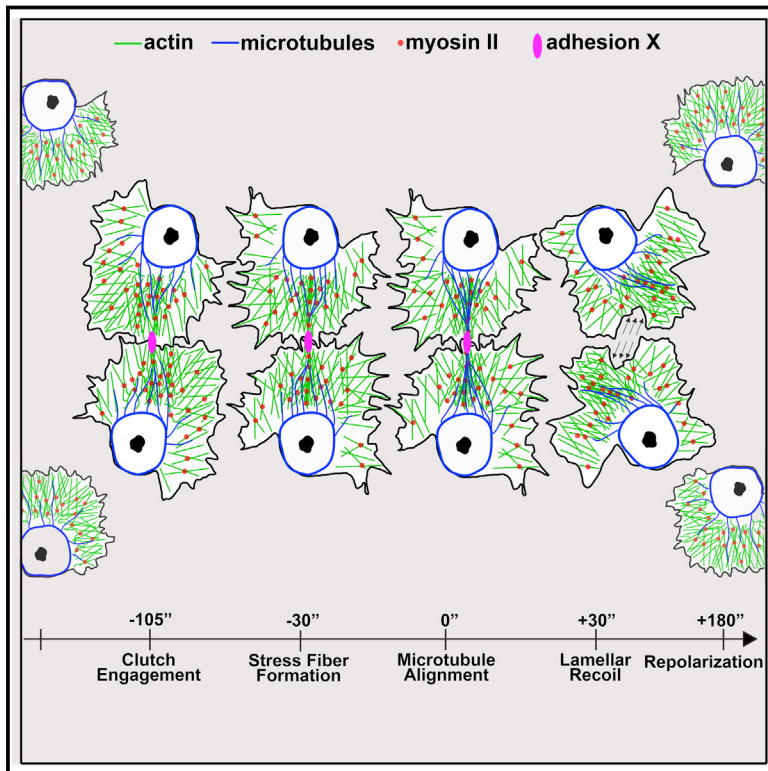
Bachmann, C., Fischer, L., Walter, U., and Reinhard, M. (1999). The EVH2 domain of the vasodilator-stimulated phosphoprotein mediates

- tetramerization, F-actin binding, and actin bundle formation. *J. Biol. Chem.* 274, 23549–23557.
- Barzik, M., Kotova, T.I., Higgs, H.N., Hazelwood, L., Hanein, D., Gertler, F.B., and Schafer, D.A. (2005). Ena/VASP proteins enhance actin polymerization in the presence of barbed end capping proteins. *J. Biol. Chem.* 280, 28653–28662.
- Bear, J.E., Svitkina, T.M., Krause, M., Schafer, D.A., Loureiro, J.J., Strasser, G.A., Maly, I.V., Chaga, O.Y., Cooper, J.A., Borisy, G.G., and Gertler, F.B. (2002). Antagonism between Ena/VASP proteins and actin filament capping regulates fibroblast motility. *Cell* 109, 509–521.
- Breitsprecher, D., Kiesewetter, A.K., Linkner, J., Urbanke, C., Resch, G.P., Small, J.V., and Faix, J. (2008). Clustering of VASP actively drives processive, WH2 domain-mediated actin filament elongation. *EMBO J.* 27, 2943–2954.
- Breitsprecher, D., Kiesewetter, A.K., Linkner, J., Vinzenz, M., Stradal, T.E.B., Small, J.V., Curth, U., Dickinson, R.B., and Faix, J. (2011). Molecular mechanism of Ena/VASP-mediated actin-filament elongation. *EMBO J.* 30, 456–467.
- Caldwell, J.E., Heiss, S.G., Mermall, V., and Cooper, J.A. (1989). Effects of CapZ, an actin capping protein of muscle, on the polymerization of actin. *Biochemistry* 28, 8506–8514.
- Campellone, K.G., and Welch, M.D. (2010). A nucleator arms race: cellular control of actin assembly. *Nat. Rev. Mol. Cell Biol.* 11, 237–251.
- Carlier, M.-F., Husson, C., Renault, L., and Didry, D. (2011). Control of actin assembly by the WH2 domains and their multifunctional tandem repeats in Spire and Cordon-Bleu. *Int. Rev. Cell Mol. Biol.* 290, 55–85.
- Cheng, A.C., and Currie, B.J. (2005). Melioidosis: epidemiology, pathophysiology, and management. *Clin. Microbiol. Rev.* 18, 383–416.
- Cooper, J.A., and Sept, D. (2008). New insights into mechanism and regulation of actin capping protein. *Int. Rev. Cell Mol. Biol.* 267, 183–206.
- Cotter, S.E., Surana, N.K., and St Geme, J.W., 3rd. (2005). Trimeric autotransporters: a distinct subfamily of autotransporter proteins. *Trends Microbiol.* 13, 199–205.
- Dautin, N., and Bernstein, H.D. (2007). Protein secretion in gram-negative bacteria via the autotransporter pathway. *Annu. Rev. Microbiol.* 61, 89–112.
- French, C.T., Toesca, I.J., Wu, T.-H., Teslaa, T., Beaty, S.M., Wong, W., Liu, M., Schröder, I., Chiou, P.-Y., Teitell, M.A., and Miller, J.F. (2011). Dissection of the Burkholderia intracellular life cycle using a photothermal nanoblade. *Proc. Natl. Acad. Sci. USA* 108, 12095–12100.
- Galyov, E.E., Brett, P.J., and DeShazer, D. (2010). Molecular insights into Burkholderia pseudomallei and Burkholderia mallei pathogenesis. *Annu. Rev. Microbiol.* 64, 495–517.
- Godoy, D., Randle, G., Simpson, A.J., Aanensen, D.M., Pitt, T.L., Kinoshita, R., and Spratt, B.G. (2003). Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, Burkholderia pseudomallei and Burkholderia mallei. *J. Clin. Microbiol.* 41, 2068–2079.
- Goldberg, M.B. (2001). Actin-based motility of intracellular microbial pathogens. *Microbiol. Mol. Biol. Rev.* 65, 595–626.
- Haglund, C.M., Choe, J.E., Skau, C.T., Kovar, D.R., and Welch, M.D. (2010). Rickettsia Sca2 is a bacterial formin-like mediator of actin-based motility. *Nat. Cell Biol.* 12, 1057–1063.
- Hansen, S.D., and Mullins, R.D. (2010). VASP is a processive actin polymerase that requires monomeric actin for barbed end association. *J. Cell Biol.* 191, 571–584.
- Haraga, A., West, T.E., Brittnacher, M.J., Skerrett, S.J., and Miller, S.I. (2008). Burkholderia thailandensis as a model system for the study of the virulence-associated type III secretion system of Burkholderia pseudomallei. *Infect. Immun.* 76, 5402–5411.
- Harris, E.S., Li, F., and Higgs, H.N. (2004). The mouse formin, FRLalpha, slows actin filament barbed end elongation, competes with capping protein, accelerates polymerization from monomers, and severs filaments. *J. Biol. Chem.* 279, 20076–20087.
- Hartmann, M.D., Grin, I., Dunin-Horkawicz, S., Deiss, S., Linke, D., Lupas, A.N., and Hernandez Alvarez, B. (2012). Complete fiber structures of complex trimeric autotransporter adhesins conserved in enterobacteria. *Proc. Natl. Acad. Sci. USA* 109, 20907–20912.
- Jewett, T.J., Fischer, E.R., Mead, D.J., and Hackstadt, T. (2006). Chlamydial TARP is a bacterial nucleator of actin. *Proc. Natl. Acad. Sci. USA* 103, 15599–15604.
- Kespichayawattana, W., Rattanachetkul, S., Wanun, T., Utainscharoen, P., and Sirisinha, S. (2000). Burkholderia pseudomallei induces cell fusion and actin-associated membrane protrusion: a possible mechanism for cell-to-cell spreading. *Infect. Immun.* 68, 5377–5384.
- Kleba, B., Clark, T.R., Lutter, E.I., Ellison, D.W., and Hackstadt, T. (2010). Disruption of the Rickettsia rickettsii Sca2 autotransporter inhibits actin-based motility. *Infect. Immun.* 78, 2240–2247.
- Kovar, D.R., Harris, E.S., Mahaffy, R., Higgs, H.N., and Pollard, T.D. (2006). Control of the assembly of ATP- and ADP-actin by formins and profilin. *Cell* 124, 423–435.
- Krause, M., Dent, E.W., Bear, J.E., Loureiro, J.J., and Gertler, F.B. (2003). Ena/VASP proteins: regulators of the actin cytoskeleton and cell migration. *Annu. Rev. Cell Dev. Biol.* 19, 541–564.
- Kühnel, K., Jarchau, T., Wolf, E., Schlichting, I., Walter, U., Wittinghofer, A., and Strelkov, S.V. (2004). The VASP tetramerization domain is a right-handed coiled coil based on a 15-residue repeat. *Proc. Natl. Acad. Sci. USA* 101, 17027–17032.
- Liverman, A.D., Cheng, H.-C., Trosky, J.E., Leung, D.W., Yarbrough, M.L., Burdette, D.L., Rosen, M.K., and Orth, K. (2007). Arp2/3-independent assembly of actin by Vibrio type III effector VopL. *Proc. Natl. Acad. Sci. USA* 104, 17117–17122.
- Mack, D., Heesemann, J., and Laufs, R. (1994). Characterization of different oligomeric species of the Yersinia enterocolitica outer membrane protein YadA. *Med. Microbiol. Immunol. (Berl.)* 183, 217–227.
- Madasu, Y., Suarez, C., Kast, D.J., Kovar, D.R., and Dominguez, R. (2013). Rickettsia Sca2 has evolved formin-like activity through a different molecular mechanism. *Proc. Natl. Acad. Sci. USA* 110, E2677–E2686.
- Meng, G., Surana, N.K., St Geme, J.W., 3rd, and Waksman, G. (2006). Structure of the outer membrane translocator domain of the Haemophilus influenzae Hia trimeric autotransporter. *EMBO J.* 25, 2297–2304.
- Namgoong, S., Boczkowska, M., Glista, M.J., Winkelman, J.D., Rebowski, G., Kovar, D.R., and Dominguez, R. (2011). Mechanism of actin filament nucleation by Vibrio VopL and implications for tandem W domain nucleation. *Nat. Struct. Mol. Biol.* 18, 1060–1067.
- Nierman, W.C., DeShazer, D., Kim, H.S., Tettelin, H., Nelson, K.E., Feldblyum, T., Ulrich, R.L., Ronning, C.M., Brinkac, L.M., Daugherty, S.C., et al. (2004). Structural flexibility in the Burkholderia mallei genome. *Proc. Natl. Acad. Sci. USA* 101, 14246–14251.
- Otomo, T., Tomchick, D.R., Otomo, C., Panchal, S.C., Machius, M., and Rosen, M.K. (2005). Structural basis of actin filament nucleation and processive capping by a formin homology 2 domain. *Nature* 433, 488–494.
- Paul, A.S., and Pollard, T.D. (2009). Review of the mechanism of processive actin filament elongation by formins. *Cell Motil. Cytoskeleton* 66, 606–617.
- Pernier, J., Orban, J., Avvaru, B.S., Jégou, A., Romet-Lemonne, G., Guichard, B., and Carlier, M.-F. (2013). Dimeric WH2 domains in Vibrio VopF promote actin filament barbed-end uncapping and assisted elongation. *Nat. Struct. Mol. Biol.* 20, 1069–1076.
- Pruyne, D., Evangelista, M., Yang, C., Bi, E., Zigmund, S., Bretscher, A., and Boone, C. (2002). Role of formins in actin assembly: nucleation and barbed-end association. *Science* 297, 612–615.
- Reed, S.C., Lamason, R.L., Risca, V.I., Abernathy, E., and Welch, M.D. (2014). Rickettsia actin-based motility occurs in distinct phases mediated by different actin nucleators. *Curr. Biol.* 24, 98–103.
- Samarin, S., Romero, S., Kocks, C., Didry, D., Pantaloni, D., and Carlier, M.-F. (2003). How VASP enhances actin-based motility. *J. Cell Biol.* 163, 131–142.
- Schafer, D.A., Jennings, P.B., and Cooper, J.A. (1996). Dynamics of capping protein and actin assembly in vitro: uncapping barbed ends by polyphosphoinositides. *J. Cell Biol.* 135, 169–179.

- Schell, M.A., Ulrich, R.L., Ribot, W.J., Brueggemann, E.E., Hines, H.B., Chen, D., Lipscomb, L., Kim, H.S., Mrázek, J., Nierman, W.C., and Deshazer, D. (2007). Type VI secretion is a major virulence determinant in *Burkholderia mallei*. *Mol. Microbiol.* **64**, 1466–1485.
- Schirenbeck, A., Arasada, R., Bretschneider, T., Stradal, T.E.B., Schleicher, M., and Faix, J. (2006). The bundling activity of vasodilator-stimulated phosphoprotein is required for filopodium formation. *Proc. Natl. Acad. Sci. USA* **103**, 7694–7699.
- Schwarz, S., Singh, P., Robertson, J.D., LeRoux, M., Skerrett, S.J., Goodlett, D.R., West, T.E., and Mougous, J.D. (2014). VgrG-5 is a *Burkholderia* type VI secretion system-exported protein required for multinucleated giant cell formation and virulence. *Infect. Immun.* **82**, 1445–1452.
- Sitthidet, C., Stevens, J.M., Field, T.R., Layton, A.N., Korbsrisate, S., and Stevens, M.P. (2010). Actin-based motility of *Burkholderia thailandensis* requires a central acidic domain of BimA that recruits and activates the cellular Arp2/3 complex. *J. Bacteriol.* **192**, 5249–5252.
- Sitthidet, C., Korbsrisate, S., Layton, A.N., Field, T.R., Stevens, M.P., and Stevens, J.M. (2011). Identification of motifs of *Burkholderia pseudomallei* BimA required for intracellular motility, actin binding, and actin polymerization. *J. Bacteriol.* **193**, 1901–1910.
- Stevens, J.M., Ulrich, R.L., Taylor, L.A., Wood, M.W., Deshazer, D., Stevens, M.P., and Galyov, E.E. (2005a). Actin-binding proteins from *Burkholderia mallei* and *Burkholderia thailandensis* can functionally compensate for the actin-based motility defect of a *Burkholderia pseudomallei* bimA mutant. *J. Bacteriol.* **187**, 7857–7862.
- Stevens, M.P., Stevens, J.M., Jeng, R.L., Taylor, L.A., Wood, M.W., Hawes, P., Monaghan, P., Welch, M.D., and Galyov, E.E. (2005b). Identification of a bacterial factor required for actin-based motility of *Burkholderia pseudomallei*. *Mol. Microbiol.* **56**, 40–53.
- Szczesny, P., and Lupas, A. (2008). Domain annotation of trimeric autotransporter adhesins—daTAA. *Bioinformatics* **24**, 1251–1256.
- Tam, V.C., Serruto, D., Dziejman, M., Brieher, W., and Mekalanos, J.J. (2007). A type III secretion system in *Vibrio cholerae* translocates a formin/spire hybrid-like actin nucleator to promote intestinal colonization. *Cell Host Microbe* **1**, 95–107.
- Toesca, I.J., French, C.T., and Miller, J.F. (2014). The Type VI secretion system spike protein VgrG5 mediates membrane fusion during intercellular spread by *pseudomallei* group *Burkholderia* species. *Infect. Immun.* **82**, 1436–1444.
- Truong, D., Copeland, J.W., and Brumell, J.H. (2014). Bacterial subversion of host cytoskeletal machinery: hijacking formins and the Arp2/3 complex. *Bioessays* **36**, 687–696.
- Welch, M.D., and Way, M. (2013). Arp2/3-mediated actin-based motility: a tail of pathogen abuse. *Cell Host Microbe* **14**, 242–255.
- Welch, M.D., Rosenblatt, J., Skoble, J., Portnoy, D.A., and Mitchison, T.J. (1998). Interaction of human Arp2/3 complex and the *Listeria monocytogenes* ActA protein in actin filament nucleation. *Science* **281**, 105–108.
- West, T.E., Frevert, C.W., Liggitt, H.D., and Skerrett, S.J. (2008). Inhalation of *Burkholderia thailandensis* results in lethal necrotizing pneumonia in mice: a surrogate model for pneumonic melioidosis. *Trans. R. Soc. Trop. Med. Hyg.* **102** (Suppl 1), S119–S126.
- Wilkinson, L. (1981). Glanders: medicine and veterinary medicine in common pursuit of a contagious disease. *Med. Hist.* **25**, 363–384.
- Winkelman, J.D., Bilancia, C.G., Peifer, M., and Kovar, D.R. (2014). Ena/VASP Enabled is a highly processive actin polymerase tailored to self-assemble parallel-bundled F-actin networks with Fascin. *Proc. Natl. Acad. Sci. USA* **111**, 4121–4126.
- Yu, B., Cheng, H.-C., Brautigam, C.A., Tomchick, D.R., and Rosen, M.K. (2011). Mechanism of actin filament nucleation by the bacterial effector VopL. *Nat. Struct. Mol. Biol.* **18**, 1068–1074.

# Inter-Cellular Forces Orchestrate Contact Inhibition of Locomotion

## Graphical Abstract



## Authors

John R. Davis, Andrei Luchici, ..., Mark Miodownik, Brian M. Stramer

## Correspondence

m.miodownik@ucl.ac.uk (M.M.),  
brian.m.stramer@kcl.ac.uk (B.M.S.)

## In Brief

Contact inhibition of locomotion in *Drosophila* macrophages involves mechanical coupling of colliding actin networks. The resultant buildup in lamellar tension orchestrates the response and choreographs cell repulsion.

## Highlights

- CIL in embryonic dispersing hemocytes requires choreographed changes in motility
- Tracking actin flow in vivo reveals synchronous changes in colliding actin networks
- An inter-cellular actin-clutch between colliding cells induces lamellar tension
- Tension buildup and release orchestrate the CIL response





# Inter-Cellular Forces Orchestrate Contact Inhibition of Locomotion

John R. Davis,<sup>1,5</sup> Andrei Luchici,<sup>1,2,5</sup> Fuad Mosis,<sup>1</sup> James Thackery,<sup>1</sup> Jesus A. Salazar,<sup>1</sup> Yanlan Mao,<sup>3</sup> Graham A. Dunn,<sup>1</sup> Timo Betz,<sup>4</sup> Mark Miodownik,<sup>2,\*</sup> and Brian M. Stramer<sup>1,\*</sup>

<sup>1</sup>Randall Division of Cell and Molecular Biophysics, King's College London, London SE1 1UL, UK

<sup>2</sup>Department of Mechanical Engineering, University College London, London WC2R 2LS, UK

<sup>3</sup>Laboratory for Molecular Cell Biology, University College London, London WC1E 6BT, UK

<sup>4</sup>Centre de Recherche, Institut Curie, Paris, UMR168, France

<sup>5</sup>Co-first author

\*Correspondence: [m.miodownik@ucl.ac.uk](mailto:m.miodownik@ucl.ac.uk) (M.M.), [brian.m.stramer@kcl.ac.uk](mailto:brian.m.stramer@kcl.ac.uk) (B.M.S.)

<http://dx.doi.org/10.1016/j.cell.2015.02.015>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## SUMMARY

Contact inhibition of locomotion (CIL) is a multifaceted process that causes many cell types to repel each other upon collision. During development, this seemingly uncoordinated reaction is a critical driver of cellular dispersion within embryonic tissues. Here, we show that *Drosophila* hemocytes require a precisely orchestrated CIL response for their developmental dispersal. Hemocyte collision and subsequent repulsion involves a stereotyped sequence of kinematic stages that are modulated by global changes in cytoskeletal dynamics. Tracking actin retrograde flow within hemocytes *in vivo* reveals synchronous reorganization of colliding actin networks through engagement of an inter-cellular adhesion. This inter-cellular actin-clutch leads to a subsequent build-up in lamellar tension, triggering the development of a transient stress fiber, which orchestrates cellular repulsion. Our findings reveal that the physical coupling of the flowing actin networks during CIL acts as a mechanotransducer, allowing cells to haptically sense each other and coordinate their behaviors.

## INTRODUCTION

Contact inhibition of locomotion (CIL), which is a cessation of forward movement upon migratory collision, is a process common to many cell types (Abercrombie and Heaysman, 1953; Astin et al., 2010; Dunn and Paddock, 1982; Gloushankova et al., 1998) that has recently been revealed to behave as a migratory cue for developmentally dispersing populations of cells during embryogenesis (Carmona-Fontaine et al., 2008; Davis et al., 2012; Stramer et al., 2010; Villar-Cerviño et al., 2013). This multifaceted phenomenon requires cells to specifically recognize each other, modulate their migratory capacity, and depending on the cell-type, subsequently repolarize. As a result of this complexity, the mechanisms behind CIL are largely unknown, and it is additionally unclear how these various behav-

iors during the process are integrated to induce a seamless response.

A range of inter-cellular adhesions and intracellular signaling pathways are postulated to be involved in CIL (e.g., Eph-ephrin [Astin et al., 2010], small GTPases [Carmona-Fontaine et al., 2008], planar cell polarity pathway [Carmona-Fontaine et al., 2008], and cell-cell adhesion [Gloushankova et al., 1998]). However, it is unclear exactly how these various signals feed into the cytoskeletal machinery to control the response. More crucially, nothing is known about the actin dynamics involved in CIL. As a central aspect of CIL is a rapid change in migration, it is clear that to understand the mechanisms behind this phenomenon it will be crucial to elucidate the dynamics of the actin network during the response.

During cell migration, the actin network provides the propulsion that allows a cell to generate movement. The actin cytoskeleton within the lamella of a migrating cell is in a constant state of retrograde flow. Actin polymerizes at the leading edge, which pushes the cell membrane forward. Subsequently, the force of polymerization against the membrane along with Myosin II driven contraction drives retrograde movement of the actin network; it is this treadmill that generates the forces behind cell motility. When a cell moves, cell-matrix receptors, such as integrins, become engaged and bind to the extracellular matrix. Integrin activation leads to a slowing of the actin flow at this integrin-based point of friction, and the force of the moving actin network is then transformed into extracellular traction stresses (Gardel et al., 2008). This integrin-dependent actin-clutch, and the resultant inverse correlation between actin flow and traction force, is hypothesized to be involved in the movement of numerous cell types.

We have been exploiting the embryonic migration of *Drosophila* macrophages (hemocytes) to understand the regulatory mechanisms of CIL and the function of this process during embryogenesis (Davis et al., 2012; Stramer et al., 2010). These cells develop from the head mesoderm and disperse throughout the *Drosophila* embryo taking defined migratory routes. One of these routes occurs just beneath the epithelium along the ventral surface where their superficial location in the embryo allows them to be imaged live at high spatio-temporal resolution approaching what can be achieved from cells in culture. This has revealed that hemocytes spread out to form an evenly

distributed pattern beneath the ventral surface within a thin acellular cavity (the hemocoel) (Stramer et al., 2010). We previously developed a mathematical model of hemocyte dispersal, and computer simulations revealed that this uniform cell spacing may be driven by contact inhibition (Davis et al., 2012). Indeed, a similar analysis of Cajal Retzius cell migration in the cerebral cortex showed an identical requirement for CIL in their dispersion (Villar-Cerviño et al., 2013), suggesting that CIL is a conserved mechanism capable of generating tiled cellular arrays.

Here, we show that hemocyte developmental dispersal requires precise contact inhibition dynamics. Quantification of changes in speed and direction during cellular collisions reveals that their CIL response is not stochastic but involves distinct kinematic stages that are synchronized between colliding partners. We also show that this choreographed movement involves a coordinated change in actin dynamics. Tracking actin flow within hemocytes *in vivo* reveals a physical coupling of the colliding actin networks through engagement of a transient inter-cellular adhesion. It is this “inter-cellular actin-clutch” and the coordinated build-up and release of lamellar tension in colliding cells that orchestrates their behaviors, allowing CIL to behave as an instructive migratory cue.

## RESULTS

### The Kinematic Steps of the CIL Response Are Synchronized in Colliding Hemocytes

Hemocytes disperse evenly within the ventral hemocoel during *Drosophila* embryogenesis and their CIL dynamics can be precisely analyzed during this process (Figure 1A; Movie S1) (Davis et al., 2012; Stramer et al., 2010). To elucidate the migratory phases of CIL, we first analyzed the changes in acceleration throughout the response with reference to the time of microtubule alignment between colliding hemocytes (Figure 1B; Movie S1), which we previously revealed is a hallmark of CIL that is associated with a change in hemocyte motility (Stramer et al., 2010). Our data revealed that there is a back acceleration upon microtubule alignment, signifying that cells were slowing down and/or changing direction (Figure 1C) (Davis et al., 2012). This was significant when calculated at either 60- or 20-s intervals (Figures 1C and S1A) highlighting that the time of microtubule alignment is correlated with a sudden change of motion during CIL.

The time of microtubule alignment allowed us to temporally register collisions and extend the time course of the acceleration analysis. We observed that 120 s before microtubule alignment there was a sudden forward acceleration, and 180 s after, an additional back acceleration event (Figure 1C). To determine whether these accelerations were due to changes in cell speed and/or direction, we quantified the internuclear distance of colliding cells during the CIL time course. Two minutes prior to microtubule alignment, the graph of internuclear distance over time revealed a sudden increase in slope (Figure 1D). This suggested that the cell speed increased, which was confirmed by analyzing the nuclear displacement rates (Figure 1E). Immediately upon microtubule alignment, the speed reduced (Figure 1E), which explained the sudden back acceleration (Figure S1A), and ~120 s later the nuclei moved apart (Figure 1D). Analysis of the

SD of the internuclear distance over time also highlighted these stages by showing an abrupt decrease in variance as cells progressed from one phase to the next, suggesting that these stages were differentially regulated (Figure S1B). These distinct phases were also visualized by calculating the average velocity vector of left and right colliding cells (Movie S1), which additionally revealed the coordinated behavior of hemocytes during CIL.

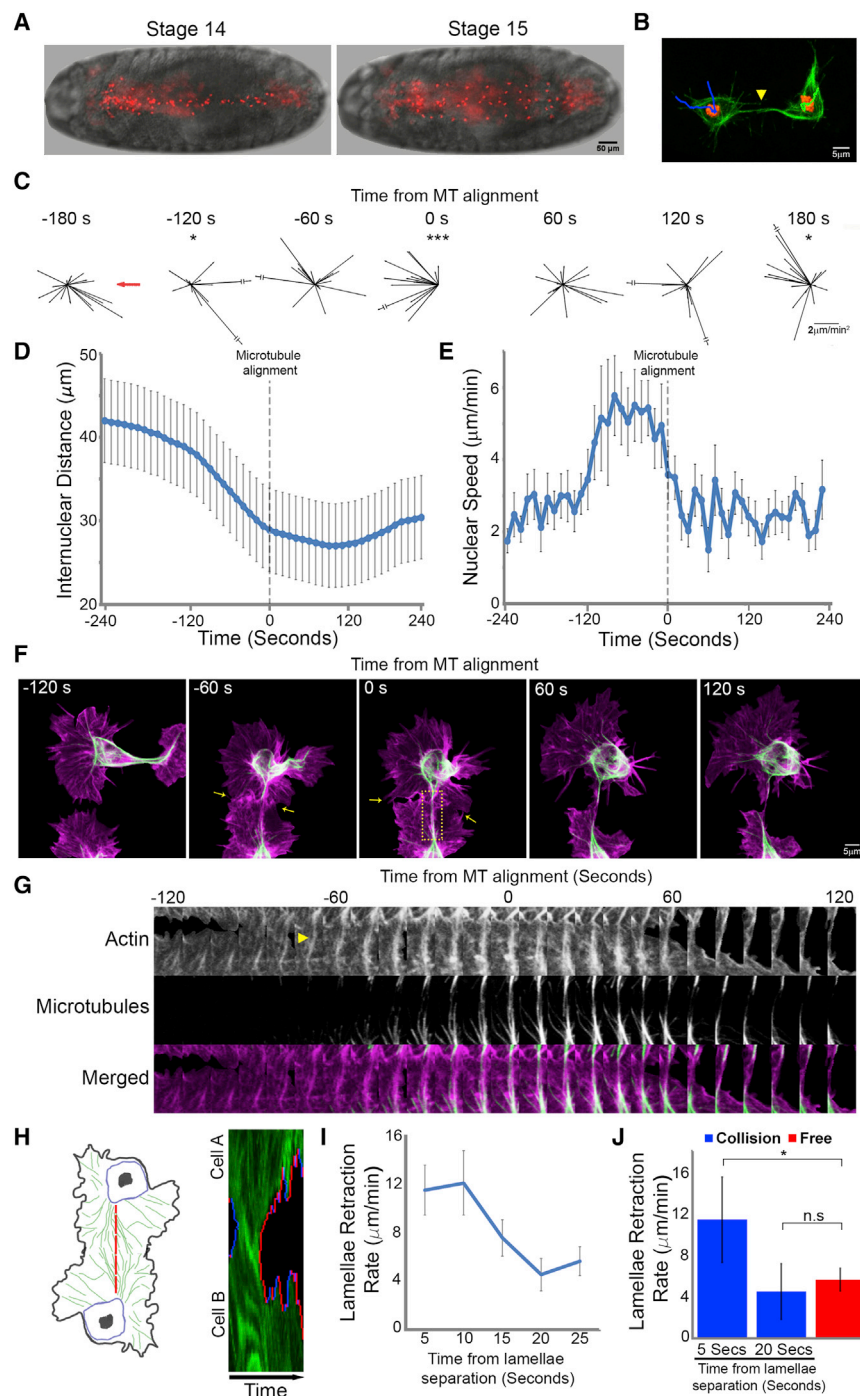
As we previously revealed that the lamellae of colliding cells overlap before microtubule alignment (Stramer et al., 2010), we hypothesized that this initial interaction was instigating the kinematic changes. Indeed, lamellae of colliding cells made contact  $105 \pm 22$  s prior to microtubule alignment (Figures 1F and 1G; Movie S1), coinciding with the forward acceleration phase of CIL. Furthermore, an actin fiber developed after lamellae contact that connected the colliding cells, which microtubules subsequently utilized as guides during alignment (Figure 1G). It is important to note that the formation of this actin fiber, and the subsequent repulsion, was not observed when a hemocyte contacted the rear of another cell (i.e., not the lamella), or collided with the lamella of a static cell (Movie S1). This analysis highlights that an interaction between lamellar actin networks of migrating cells is initiating the CIL response.

Analysis of the separation phase of CIL also revealed a synchronous response between colliding cells. Kymography of lamellar retraction revealed that colliding partners simultaneously retracted their lamellae at two to three times the speed of retraction events of freely moving cells (Figures 1H–1J). This retraction event initiated  $32 \pm 22$  s after microtubule alignment, which coincided with the initiation of movement away from the colliding partner. The retraction of lamellae occurred prior to the development of new protrusions away from the colliding partner, suggesting that rapid lamellar retraction initiates cell repolarization (Figures S1C and S1D). Different cell-types exhibit distinct CIL behaviors, which are classified as either type 1 (involving contact-induced lamellar contraction) or type 2 (inhibition of locomotion without contraction) responses (Stramer et al., 2013). Our analysis suggests that hemocytes undergo a classical type 1 response similar in description to chick heart fibroblasts, which were also observed to exhibit sudden lamellar recoil (Abercrombie and Heaysman, 1953).

### The Actin Cytoskeleton Rapidly Reorganizes in Colliding Partners during CIL

To understand how the actin networks were mediating the response, we analyzed the actin retrograde flow dynamics during CIL. Time-lapse movies of freely moving hemocytes, labeled with the actin probe, LifeAct-GFP, during their developmental dispersal revealed a highly dynamic actin network within their lamellae (Movie S2). We adapted a fluorescent pseudo-speckle tracking technique (Betz et al., 2009) to quantify the precise speed and direction changes of actin flow within hemocytes *in vivo* (Figure S2A; Movie S2). This revealed that freely moving hemocytes *in vivo* have a mean actin flow rate of  $3.2 \pm 1.8$   $\mu\text{m}/\text{min}$ , which is similar to growth cones *in vitro* (Betz et al., 2009).

Live imaging of collisions revealed significant reorganization of the actin networks during the response (Figure 2A; Movie S2). This reorganization coincided with the development of an actin fiber, which ran perpendicular to the leading edge, linking the



**Figure 1. Hemocyte Contact Inhibition Involves Multiple Stages that Are Synchronous and Coordinated in Colliding Partners**

(A) Dispersal of hemocytes labeled with a nuclear marker (red) beneath the ventral surface of a *Drosophila* embryo (bright-field) at developmental stages 14 and 15.

(B) Automatic tracking of nuclei (red) of colliding hemocytes while also registering collisions with microtubules (green). Time point of microtubule alignment (arrowhead) allows for temporal registration of CIL events in subsequent kinematic analyses.

(C) Time course of hemocyte accelerations (black arrows) surrounding a collision event with reference to the colliding partner (red arrow). All time points show random accelerations except at  $-120$ ,  $0$ , and  $180$  s where there is a bias along the x axis.  $*p < 0.05$ ,  $***p < 0.001$ .

(D) Graph showing the internuclear distance of colliding cells during the CIL time course. Note the change in slope at  $-120$ ,  $0$ , and  $120$  s. Error bars represent SD.

(E) Graph showing nuclear speed during collisions. Note the increase in speed at  $-120$  s and the subsequent decrease upon microtubule alignment. Error bars represent SD.

(F) Time-lapse sequence of colliding hemocytes labeled with an F-actin (magenta) and a microtubule (green) probe. Arrows highlight region of lamellae overlap.

(G) Kymograph of the region highlighted in (F) showing the time course of actin fiber formation (arrowhead highlights the initial development of the actin fiber) and microtubule alignment.

(H) Kymograph of lamellar activity (red regions show lamellar retraction and blue extension) in colliding partners along the actin fiber (red dotted line in schematic). Note that retraction is simultaneous in colliding cells upon lamellae release.

(I) Quantification of the rate of lamella retraction over time. Error bars represent SEM.

(J) Quantification of lamella retraction rates at 5 and 20 s after cell separation compared with average retraction rates in freely moving cells. Error bars represent SD.  $*p < 0.05$ .

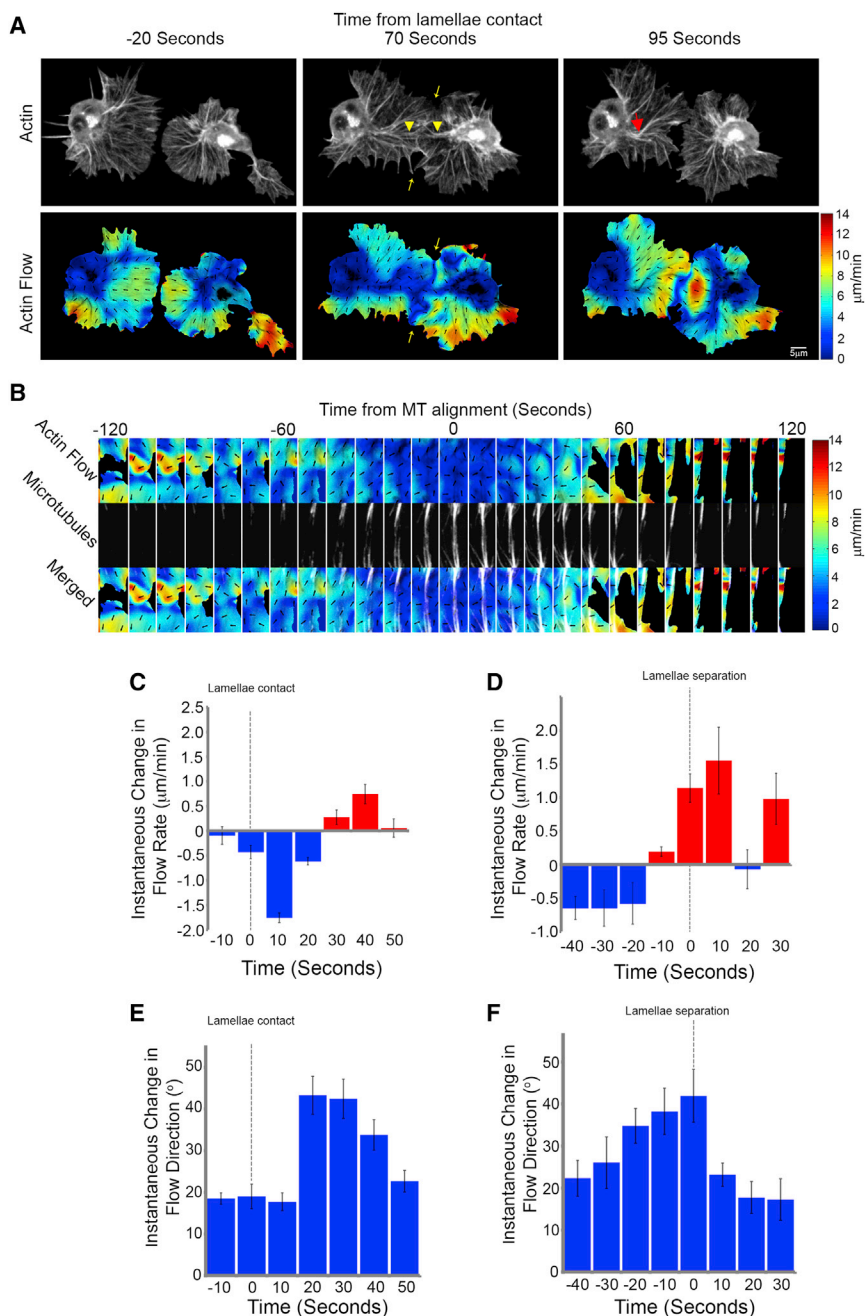
See also Figure S1 and Movie S1.

lamellae of colliding cells (Figures 2A and S2B). Pseudo-speckle microscopy of collisions highlighted a slowing of the actin flow within a corridor that colocalized with the actin fiber (Figure 2A; Movie S2) and the aligned microtubule bundle (Figure 2B; Movie S2). Quantification of the actin flow rate within the region surrounding the actin fiber revealed a decrease in magnitude during the response, which suddenly increased upon lamellae separation (Figures 2C, 2D, and S2C; Movie S2). It is interesting to

followed by an additional increase during cellular repolarization (Figures 2D and S2C). Analysis of instantaneous changes in flow direction also revealed an increase in rotation after lamellae contact (Figure 2E), which rapidly returned to levels observed in freely moving cells after lamellae separation (Figure 2F; Movie S2). The change in flow direction coincided with a movement of actin fibers within the lamella toward the nascent actin cable, which contributed to its formation (Figure S2D). Upon lamellae

note that these analyses highlight that the increase in actin flow speed occurred in two phases; there was an abrupt spike immediately upon lamellae separation lasting  $\sim 20$  s (that coincided with the duration of lamella recoil) (Figure 1I), fol-





**Figure 2. During Contact Inhibition, the Actin Network Is Rapidly Reorganized in Colliding Partners**

(A) Top panels are still images from a time-lapse movie of hemocytes containing labeled F-actin during a collision. While cells are in contact, an actin fiber develops between the cell body and the point of contact in colliding partners (arrowheads), which often deforms and breaks upon lamellar retraction (red arrow). Bottom panels highlight actin flow dynamics obtained from the pseudo-speckle analysis. Note that the decreased actin flow in the vicinity of lamellae overlap (highlighted by yellow arrows) is due to the inability of the algorithm to distinguish between the two networks.

(B) Kymograph of the region surrounding the actin fiber highlighting the actin retrograde flow dynamics and the alignment of the microtubule bundles (pseudocolored white).

(C and D) Instantaneous changes in retrograde flow rate quantified from lamellae contact (C) or lamellae separation (D).

(E and F) Instantaneous changes in retrograde flow direction quantified from lamellae contact (E) or lamellae separation (F).

For (C)–(F), error bars represent SD. See also [Figure S2](#) and [Movie S2](#).

cell-cell adhesion was responsible for the rapid and seemingly coordinated actin network changes in colliding hemocytes. As Zyxin is known to be a marker of both cell-matrix and cell-cell adhesions ([Hirata et al., 2008](#)), we expressed mCherry-Zyxin in hemocytes and colocalized Zyxin-labeled adhesions with actin during CIL. Immediately upon lamellae overlap, a punctum of Zyxin developed at the site of cell-cell contact and persisted for the duration of the response ([Figures 3A and 3B](#); [Movie S3](#)). Subsequently, the actin fiber formed immediately behind this concentration of Zyxin ([Figures 3A and 3B](#); [Movie S3](#)). We hypothesized that Zyxin foci represented transient cell-cell adhesions that modulate actin retrograde flow in a process analogous to the integrin-based actin-clutch reported in migrating cells in vitro

separation, the actin fiber deformed and was subsequently lost as the actin flow rapidly returned to its normal retrograde direction ([Figure 2A](#); [Movie S2](#)). These data highlight that CIL involves a dramatic reorganization of the lamellar actin network.

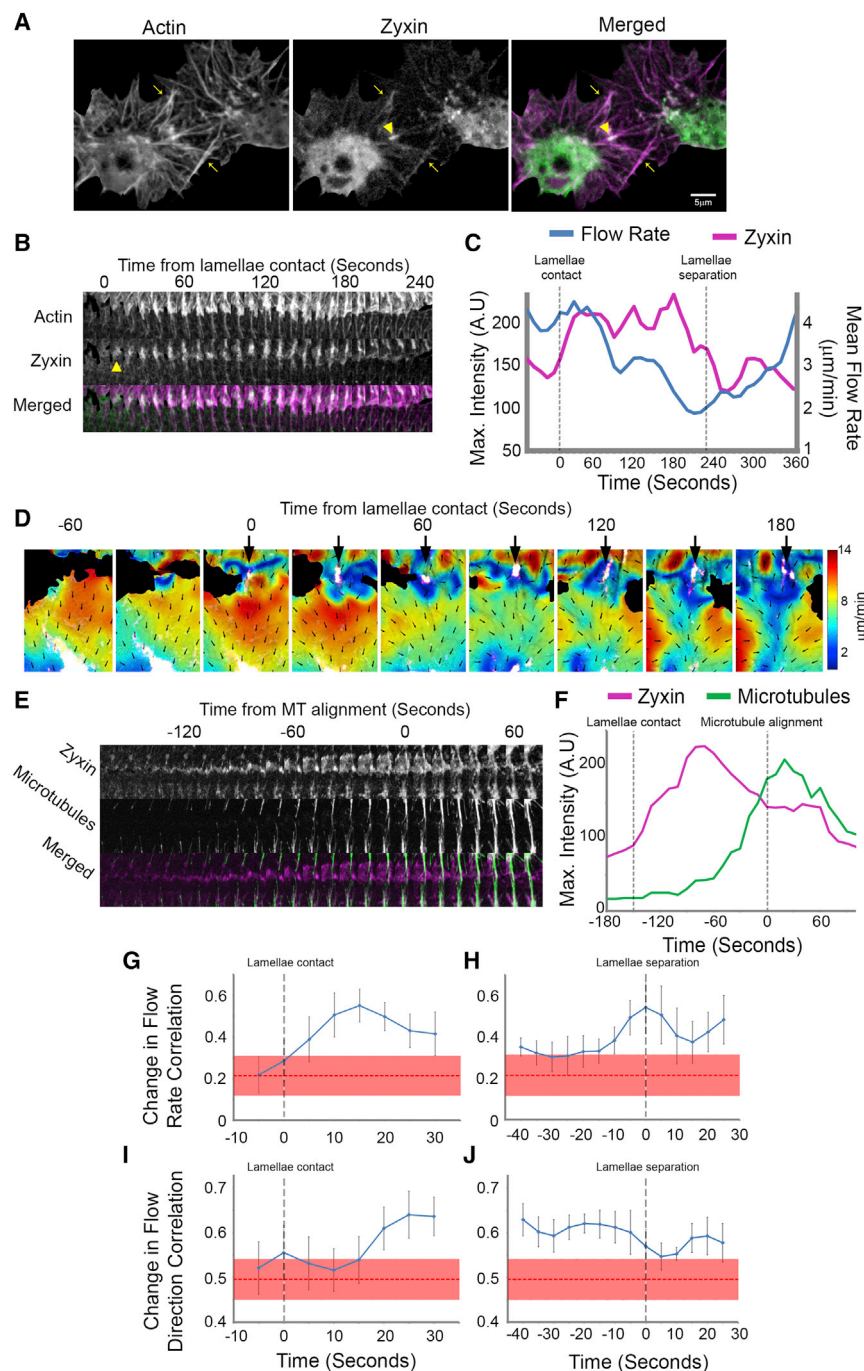
#### Development of a Transient Cell-Cell Adhesion during CIL Coincides with a Coordinated Reorganization of the Actin Network

As a number of cell types have been reported to form transient inter-cellular adhesions during CIL ([Gloushankova et al., 1998](#); [Theveneau et al., 2010](#)), we wanted to determine whether a

([Gardel et al., 2008](#)). Indeed, visualization of Zyxin while analyzing actin flow revealed that after development of the Zyxin puncta the retrograde flow rate decreased within a corridor immediately behind ([Figures 3C and 3D](#); [Movie S3](#)). Furthermore, microtubules polymerized toward this site of adhesion ([Figure 3E](#); [Movie S3](#)). Immediately upon microtubule targeting of this adhesion, Zyxin levels decreased ([Figure 3F](#)).

The development of an inter-cellular adhesion during CIL suggested that the colliding actin networks were becoming physically coupled. We therefore examined whether this coupling could lead to synchronous changes in actin retrograde





**Figure 3. Actin Network Reorganization Correlates with the Formation of a Transient Cell-Cell Adhesion**

(A) Still image of a collision between hemocytes expressing mCherry-Zyxin (green) and labeled F-actin (magenta), which highlights the inter-cellular adhesion at the point of initial contact (arrowhead). Arrows highlight region of lamellae overlap.

(B) Kymograph of Zyxin and actin dynamics in the region of the actin fiber. Note that the punctum of Zyxin forms in line with the actin fiber and persists for the duration of the time in contact (arrowhead highlights the initial formation of the punctum).

(C) Quantification of the maximum intensity of Zyxin and average actin flow rate during the collision.

(D) Analysis of actin flow dynamics in comparison with Zyxin localization (pseudocolored white). Note that the region of low retrograde flow develops in line with the inter-cellular adhesion (arrows).

(E) Kymograph of Zyxin and microtubule dynamics in the region of the actin fiber highlighting microtubule targeting of the Zyxin puncta.

(F) Maximum intensity of Zyxin and microtubules at the inter-cellular adhesion in the region highlighted in (E).

(G–J) Cross correlation of the instantaneous changes in flow rate (G and H) and flow direction (I and J) in lamellae of colliding cells. Error bars represent SEM. Red dotted lines represent the mean correlation between colliding cells immediately prior to cell-cell contact with the thickness representing the SEM.

See also [Movie S3](#).

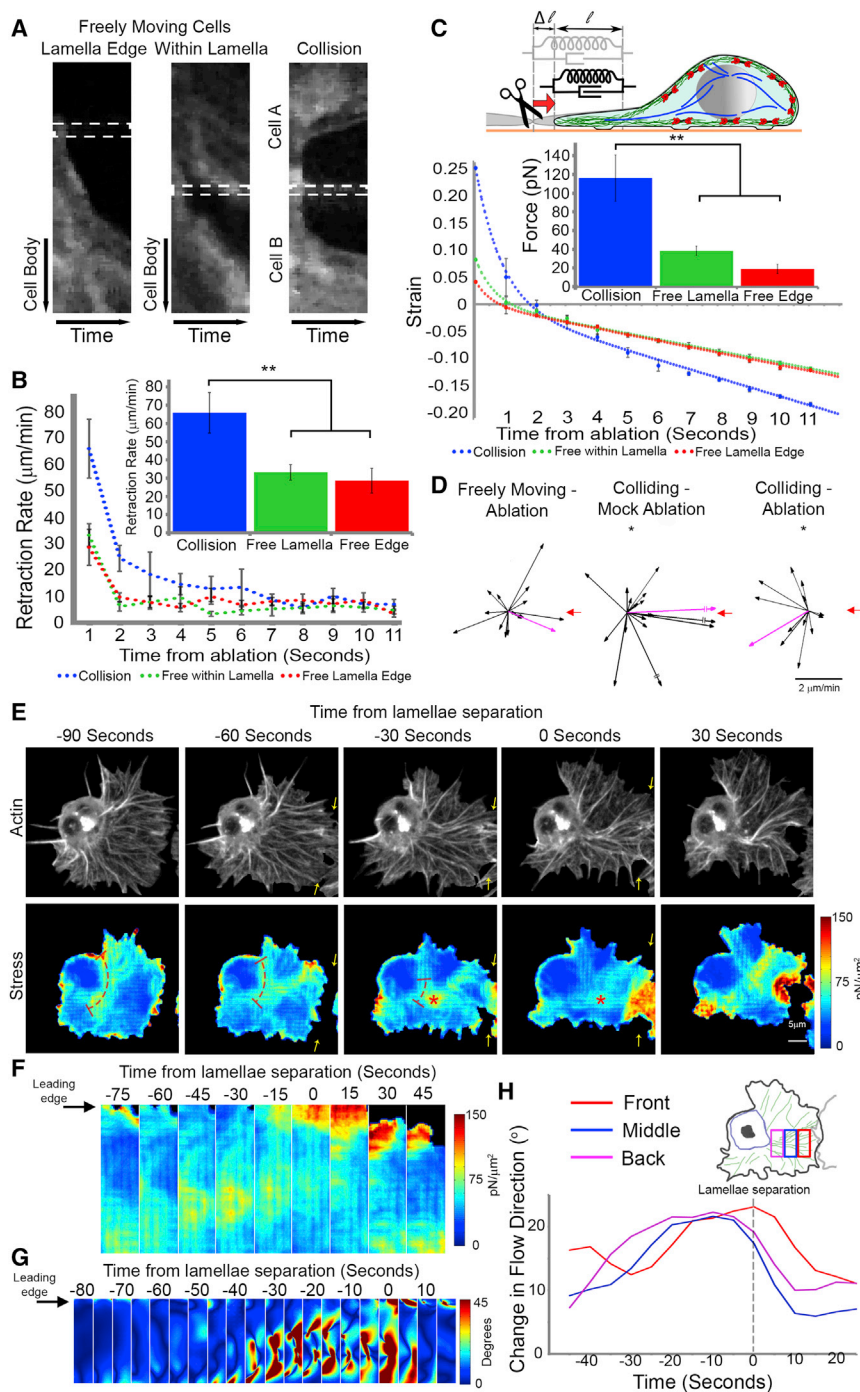
tact, which remained high until the time of separation (Figures 3I and 3J). These data reveal that, similar to the orchestrated motion of colliding cells, the actin networks are behaving in a coordinated fashion during CIL.

### Increase and Redistribution of Actin Network Stress during Cell Collision

The sudden and synchronous retraction event that occurred upon cell separation (Figure 1H) suggested that tension is developing within the actin network during CIL. We therefore analyzed lamellar tension by laser abscission of the actin cytoskeleton as the recoil rate of the

flow in colliding cells. Investigation of the correlation between instantaneous changes in flow speed in colliding partners revealed an increase immediately upon lamellae overlap, which slowly diminished as cells remained in contact (Figure 3G). Subsequently, at the time of lamellae separation there was an additional abrupt increase in the correlation of instantaneous changes in flow speed (Figure 3H; [Movie S3](#)). Similarly, correlation between the instantaneous changes in flow direction of colliding partners showed an increase ~20 s after lamellae con-

network is indicative of tension within the lamella. Ablating the leading edge or an actin fiber within the lamellae of a freely moving cell led to an initial recoil rate of  $28.6 \pm 6.8$  and  $33.2 \pm 4.3$   $\mu\text{m}/\text{min}$ , respectively. Interestingly, the recoil was unidirectional toward the cell body when the ablation was performed within the lamella (Figures 4A and 4B; [Movie S4](#)), which may be explained by a bias of myosin contraction toward the rear of the lamella (Svitkina et al., 1997; Yam et al., 2007). In contrast, cutting the region of lamella overlap across the actin fiber linking colliding



**Figure 4. Lamellar Stresses Are Increased and Redistributed during CIL**

(A) Kymographs of lamellar recoil upon laser abscission of the actin network in freely moving and colliding cells. Dotted rectangle highlights the width of the ablation region.

(B) Quantification of recoil rate over time and initial recoil rate upon laser abscission. Error bars represent SEM.  $**p < 0.01$ .

(C) Quantification of lamellae strain over time upon laser abscission and modeled forces assuming that the actin network behaves elastically over short time scales. The elastic and dissipative mechanical properties in the lamellae are modeled by an exponential decay of the strain that is overlaid onto the constant retrograde flow. Note that zero strain represents the end of the exponential decay. Assuming mechanical properties similar to previously published lamellae we can estimate the tension. Inset: Sketch illustrating the mechanical model of an elastic and dissipative element. The strain  $u$  is calculated by the ratio  $\Delta l / l$ . Error bars represent SEM.  $**p < 0.01$ .

(D) Hemocyte velocities in freely moving and colliding cells 60 s after laser abscission with respect to the ablation site (red arrow). Magenta arrow is the average direction of the population. Note that after mock ablation there was a significant forward movement of cells, while ablation of the fiber during cell collision led to a significant rearward movement.  $*p < 0.05$ .

(E) Localization of actin network stress during cell collision. Top panels: a time-lapse series of a hemocyte containing labeled F-actin undergoing a collision (adapted from Figure 2A). Bottom panels: modeled intracellular actin stresses. Note that stresses were only measured for regions of the lamella that persisted for a 40-s period as deformation history is required in the analysis. Arrows highlight region of lamellae overlap. Dotted line highlights the redistribution of stresses around the cell body and asterisks the regions of high stress that colocalize with the actin fiber.

(F) Kymograph of lamellar stresses over the region that colocalized with the actin fiber. Note the redistribution of stress from the back of the network to the front.

(G) Kymograph of the instantaneous changes in actin flow direction in the region colocalizing with the actin fiber.

(H) Quantification of the mean change in flow direction of the actin network in three regions corresponding to the back, middle, and front of the actin fiber. Note that the changes initially increase in the rear of the network. See also Movie S4.

cells led to a significantly enhanced retraction rate of  $65.8 \pm 11.1 \mu\text{m}/\text{min}$  suggesting increased tension was stored within the actin network during CIL (Figures 4A and 4B; Movie S4). We subsequently modeled the amount of force present within the actin network by assuming that the actin cytoskeleton behaves elastically over short time scales (Gardel et al., 2004). Tracking edge displacement of the lamellae upon laser abscis-

sion allowed us to measure the changes in strain of the lamellar network as it retracts. This revealed that laser abscission of the lamellae leads to an initial rapid exponential decay that is caused by the sudden release of lamellar tension, followed by a slower linear phase as the retrograde actin flow continuously pulls in the network (Figure 4C). Assuming a lamellar stiffness similar to growth cones, the measured strain upon the release of the

lamellar tension allowed us to infer the forces present within the actin network in freely moving and colliding cells. This revealed that forces in colliding lamellae were  $\sim 3$ -fold higher than freely moving cells (Figure 4C). Furthermore, release of this tension by laser abscission during cell collision was sufficient to induce a rearward movement of the cell body away from the ablation site (Figure 4D; Movie S4).

We also adapted previously developed techniques to infer the localization of stresses that drive the retrograde flow using an estimation of actin network deformation (Betz et al., 2011; Ji et al., 2008). As previous correlation of flow changes suggested that the colliding actin networks were becoming coupled (Figures 3G–3J), the model assumed that the colliding networks were behaving as a single visco-elastic material. Interestingly, this analysis revealed that there was a redistribution of stresses during the response. Prior to cell-cell contact, most of the stress was localized around the cell body similar to freely moving hemocytes. However, upon collision the stress redistributed from around the cell body to the base of the actin fiber (Figure 4E; Movie S4) where it subsequently propagated to more distal regions of the fiber (Figure 4F). After cell separation, the stress became localized again to the region around the cell body as in freely moving cells (Movie S4). This stress redistribution indicated that actin cytoskeletal changes were spreading, not from the site of lamellae contact, but from the rear of the network. Indeed, when we examined the distribution of instantaneous changes in the direction of actin flow, we observed a similar propagation from the cell rear (Figures 4G and 4H). These data confirm that lamellar tension increases during CIL and propagates from the rear of the network, suggesting that these cytoskeletal changes are not initiated by a local signal released from the site of cell contact.

### Coupling of Colliding Actin Networks Leads to the Development of a Transient Stress Fiber

As Myosin II is the major motor responsible for generating contraction within the actin cytoskeleton, we examined its dynamics during hemocyte migration. In freely moving cells, both actin and Myosin II flow in a similar retrograde fashion (Figures 5A, 5B, S3A, and S3B). However, comparable to our previously observed changes in actin flow direction during CIL (Movie S2), Myosin II flow reoriented perpendicularly toward the actin fiber (Figures 5C, 5D, and S3C; Movie S5). The actin fiber developed coincidentally with Myosin II accumulation along its length (Figures 5E and 5F) and subsequently became decorated with repeating puncta of Myosin II  $\sim 1.4 \mu\text{m}$  apart (Figure S3D), similar to stress fibers in vitro (Hotulainen and Lappalainen, 2006). Furthermore, the microtubule bundle aligned along the Myosin II decorated actin fiber (Movie S5). Analogous to the propagation of modeled stresses from the rear of the actin fiber, Myosin II intensity first increased at the back of the lamella (Figure 5G) and colocalized with the corridor of reduced retrograde flow (Figure 5H). Upon lamellae retraction, the Myosin II puncta along the actin fiber rapidly moved in a retrograde fashion with the actin network toward the cell body (Movie S5).

We also examined the localization of the formin, Diaphanous (Dia), which has previously been shown to decorate stress fibers within cells in vitro and be important for stress fiber assembly

(Nakano et al., 1999; Sandbo et al., 2013). Expression of constitutively active Dia within hemocytes led to reduced cell spreading and an accumulation of Myosin within the lamella, with periods of severe contraction of the actin network suggesting that activation of Dia can increase lamellar tension (Figure S3E; Movie S5). As *Drosophila* Dia is also known to be involved in filopodia formation (Homem and Peifer, 2009), it was unsurprising to observe wild-type Dia localized to filopodia in freely moving cells (Figures S3F and S3G). However, upon cell collision, Dia localized along the actin fiber (Figures 5I and 5J). These data suggest that the previously described actin fiber is a stress fiber-like structure that couples colliding actin networks during CIL.

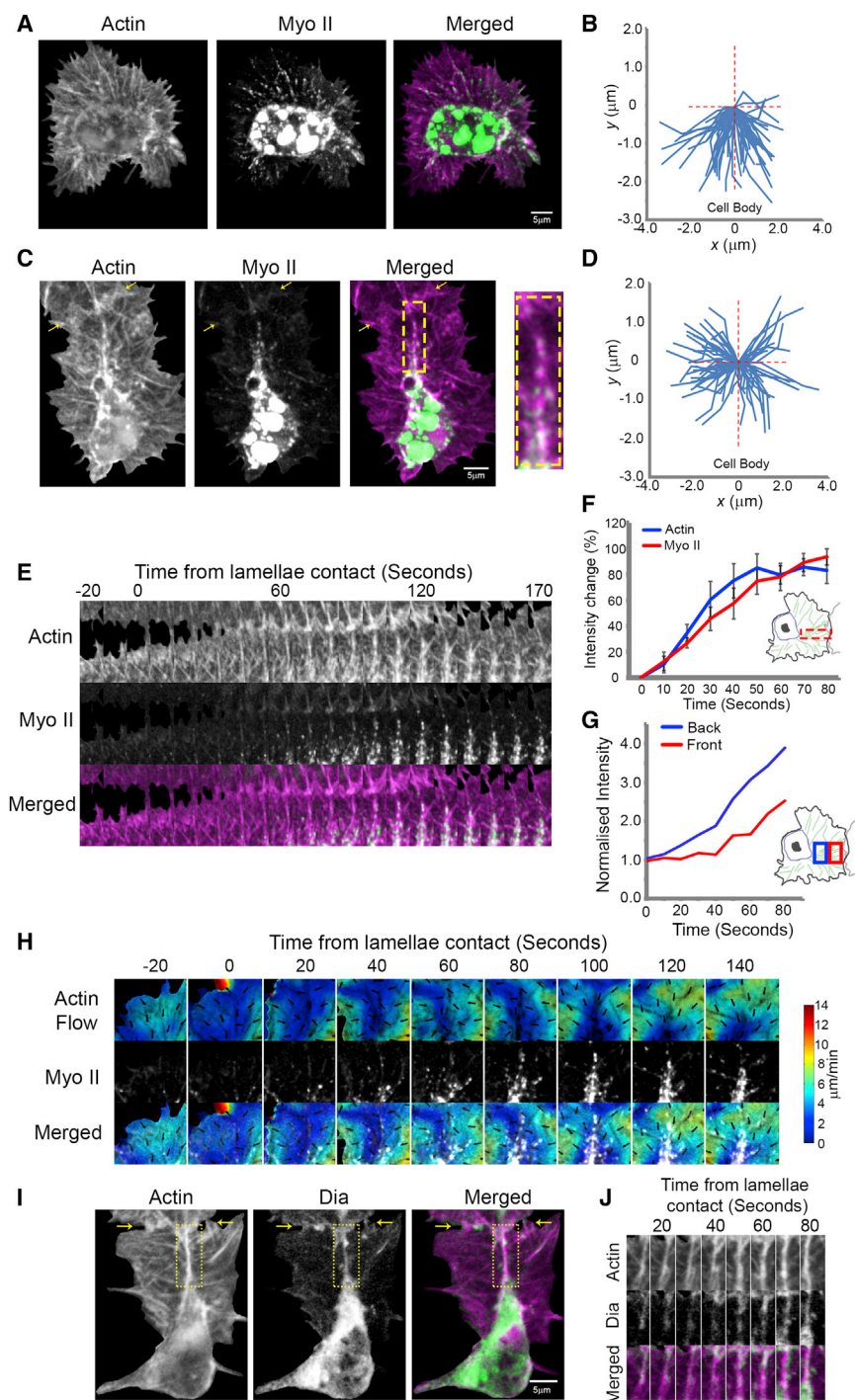
### Myosin II-Dependent Contraction and Stress Fiber Formation Are Essential for Coordinating the CIL Response in Colliding Cells

We subsequently examined the importance of lamellar contraction and stress fiber formation in regulating the CIL response. We first analyzed actin flow in hemocytes of embryos mutant for *myosin II heavy chain* (hereafter mentioned as Myosin II and called Zipper [Zip] in the fly). *Drosophila* zygotic mutant embryos (*zip*<sup>1</sup>) only begin to show defects at late stages of embryogenesis (when we also begin to image hemocyte motility) suggesting that maternal levels of the protein perdure to this stage of development (Young et al., 1993). Mutant hemocytes were initially capable of migrating from the head where they originate, but began to fail in motility soon after this stage (Figure S4A). This reveals that, in contrast to the reported dispensability of Myosin II in cell migration in 2D environments in vitro (Doyle et al., 2012), Myosin II is critical for hemocyte motility during embryogenesis. Analysis of retrograde flow in *myosin II* mutants showed that they had a significant reduction in speed ( $1.5 \pm 1.0 \mu\text{m}/\text{min}$ ) (Figures S4B and S4C; Movie S6). Expression of a GFP-tagged Myosin II specifically in hemocytes within *zip*<sup>1</sup> mutants rescued their developmental dispersal (Figure S4A) and retrograde flow rates (Figures S4B and S4C), showing that Myosin II is critical for actin flow in vivo. These data suggest that *myosin II* mutant hemocytes have a reduction in the contractility of their actin networks.

We subsequently examined the lamellar dynamics of *myosin II* mutant hemocytes during collisions. Time-lapse movies of *zip*<sup>1</sup> hemocyte collisions revealed that colliding cells failed to reorganize their actin networks (Figure 6A), or increase their lamellar retraction rates upon separation (Figures 6B and 6C; Movie S6). They also did not develop a prominent actin fiber during cell collision (Figure 6D). These defects were accompanied by a failure to cease their forward motion upon collision and an increased time in contact during the CIL response (Figures 6E and 6F). However, similar to fibroblasts in vitro (Giannone et al., 2007), loss of Myosin II in freely moving hemocytes also reduced their rate of lamellar retraction, showing that they had general defects in lamellar tension (Figure 6C).

We therefore specifically analyzed the role of the stress fiber during CIL by examining the response in *diaphanous* (*dia*<sup>5</sup>) mutants. In contrast to *zip*<sup>1</sup> hemocytes, freely moving *dia*<sup>5</sup> mutant cells showed no aberration in actin retrograde flow (Figures S5A–S5C; Movie S6), cell migration (Figures S5D and S5E), or rate of lamellar retraction (Figure 6I). However, colliding cells





**Figure 5. The Actin Fiber that Couples Colliding Cells Is a Stress Fiber-like Structure**

(A) Still image of a freely moving hemocyte containing labeled F-actin (magenta) and Myosin II (green).

(B) Quantification of Myosin II tracks in freely moving cells.

(C) Still image of a collision between hemocytes containing labeled actin and Myosin II. Note the puncta of Myosin II along the actin fiber (inset). Arrows highlight region of lamellae overlap.

(D) Quantification of Myosin II tracks for 40 s upon lamellae overlap during CIL.

(E) Kymograph of the region surrounding the actin fiber in (C) highlighting Myosin II accumulation during a collision.

(F) Quantification of the increase in actin and Myosin II intensity in the region corresponding to the actin fiber relative to values prior to lamellae contact. Error bars represent SEM.

(G) Quantification of Myosin II intensity in regions corresponding to the back versus the front of the actin fiber during CIL.

(H) Analysis of actin flow dynamics in comparison with Myosin II localization (pseudocolored white). Note that actin network reorganization precedes Myosin II accumulation along the stress fiber.

(I) Still image of a collision between hemocytes containing labeled actin (magenta) and Diaphanous (green). Arrows highlight region of lamellae overlap.

(J) Kymograph of the region surrounding the actin fiber in (I).

See also [Figure S3](#) and [Movie S5](#).

and persisted in contact for a longer duration ([Figure 6F](#)). Despite these defects, *dia*<sup>5</sup> hemocytes, similar to wild-type cells, formed a Zyxin puncta and showed some microtubule alignment during collision ([Figure S6](#)). These data reveal that preventing stress fiber formation results in an aberrant response whereby cells eventually separate, but in the absence of the sudden lamellar retraction characteristic of type 1 CIL.

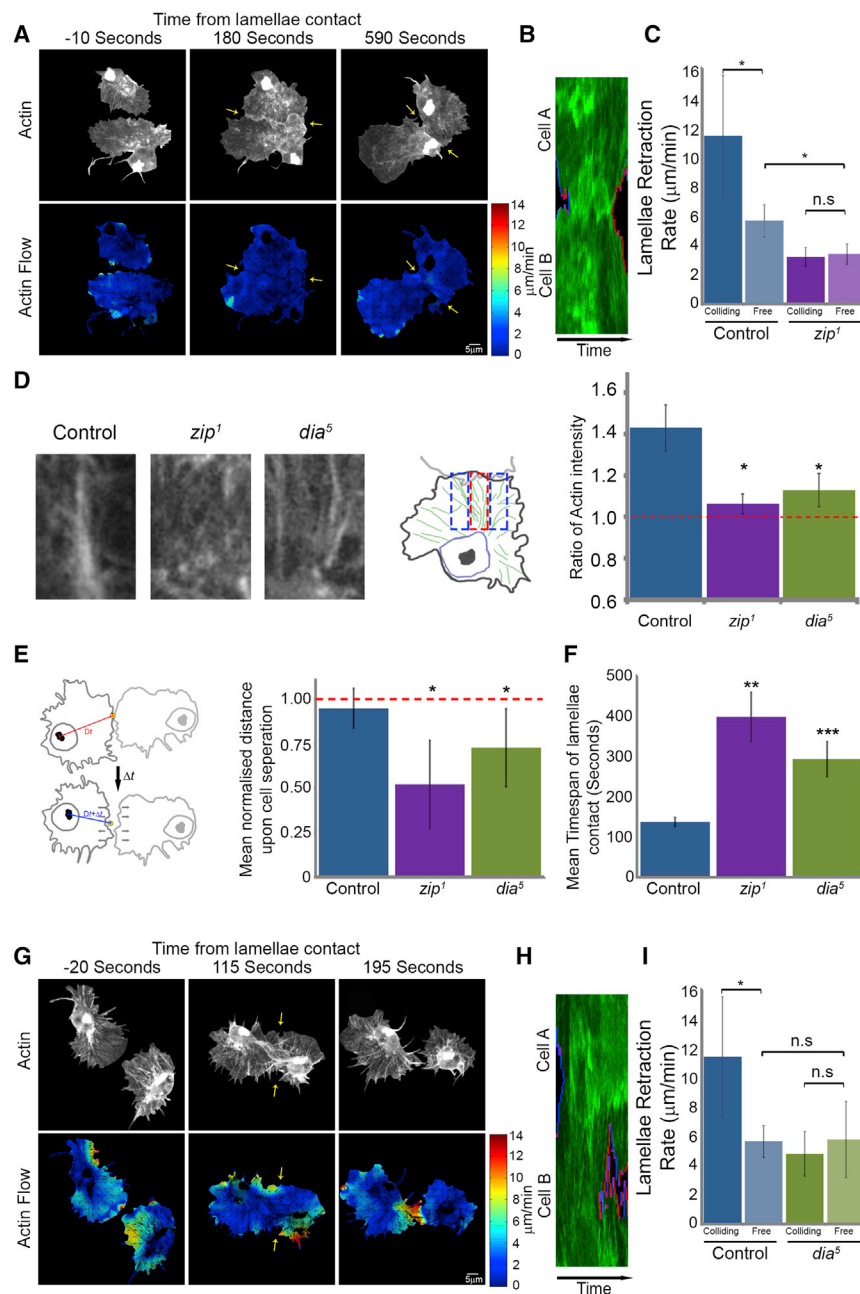
### A Coordinated CIL Response Is Essential for Hemocyte Dispersal

As *dia*<sup>5</sup> mutant hemocytes showed uncoordinated cytoskeletal dynamics during collisions ([Figures S5G](#) and [S5I](#)), we

showed a highly variable response ([Movie S6](#)) and, on average, failed to generate an actin fiber ([Figure 6D](#)) or coordinate their actin dynamics ([Figures 6G](#) and [S5F–S5I](#)). Additionally, their lamellar retraction rates upon collision were no different from freely moving cells ([Figures 6H](#) and [6I](#)). Furthermore, similar to *zip*<sup>1</sup> hemocytes, *dia*<sup>5</sup> mutant cells showed a reduced capacity to cease their forward movement upon collision ([Figure 6E](#))

wanted to determine whether this defect also led to an uncoordinated kinematic response during CIL. Analysis of the acceleration time course revealed that, in contrast to the three acceleration changes observed in wild-type hemocytes ([Figure 1C](#)), *dia*<sup>5</sup> mutant cells only showed the back acceleration upon microtubule alignment ([Figure 7A](#)). However, this back acceleration was significant only when calculated at 60-s





**Figure 6. A Stress Fiber-like Structure Is Required for a Normal CIL Response**

(A–C) Myosin II mutant (*zip*<sup>1</sup>) collisions. (A) Top panels are still images from a time-lapse movie of hemocytes containing labeled F-actin during a collision. Bottom panels are heatmaps obtained from the pseudo-speckle analysis showing no substantial changes in retrograde flow. Arrows highlight region of lamellae overlap. (B) Kymograph of lamellar activity in colliding partners in a region perpendicular to the point of cell contact (red regions highlight lamellar retraction and blue extension). (C) The speed of lamellar retraction in *myosin II* mutants was quantified at the time of separation to reveal that the retraction rate was no different to freely moving cells. Error bars represent SD. \*p < 0.05. Note that control retraction rates are from Figure 1J.

(D) Quantification of actin fiber formation in control, *zip*<sup>1</sup> and *dia*<sup>5</sup> mutant hemocytes during CIL. The graph represents the relative increase in actin intensity within the region encompassing the actin fiber (red box in schematic) with respect to the surrounding regions of the actin network (blue boxes in schematic).

(E) Quantification of the cessation of forward movement during CIL in which the mean distance between the initial point of contact and the nucleus was measured and compared to the distance at the time of cell separation. This analysis revealed that the *zip*<sup>1</sup> and *dia*<sup>5</sup> mutants failed to inhibit their forward motion in comparison to control cells. Error bars represent SD. \*p < 0.05.

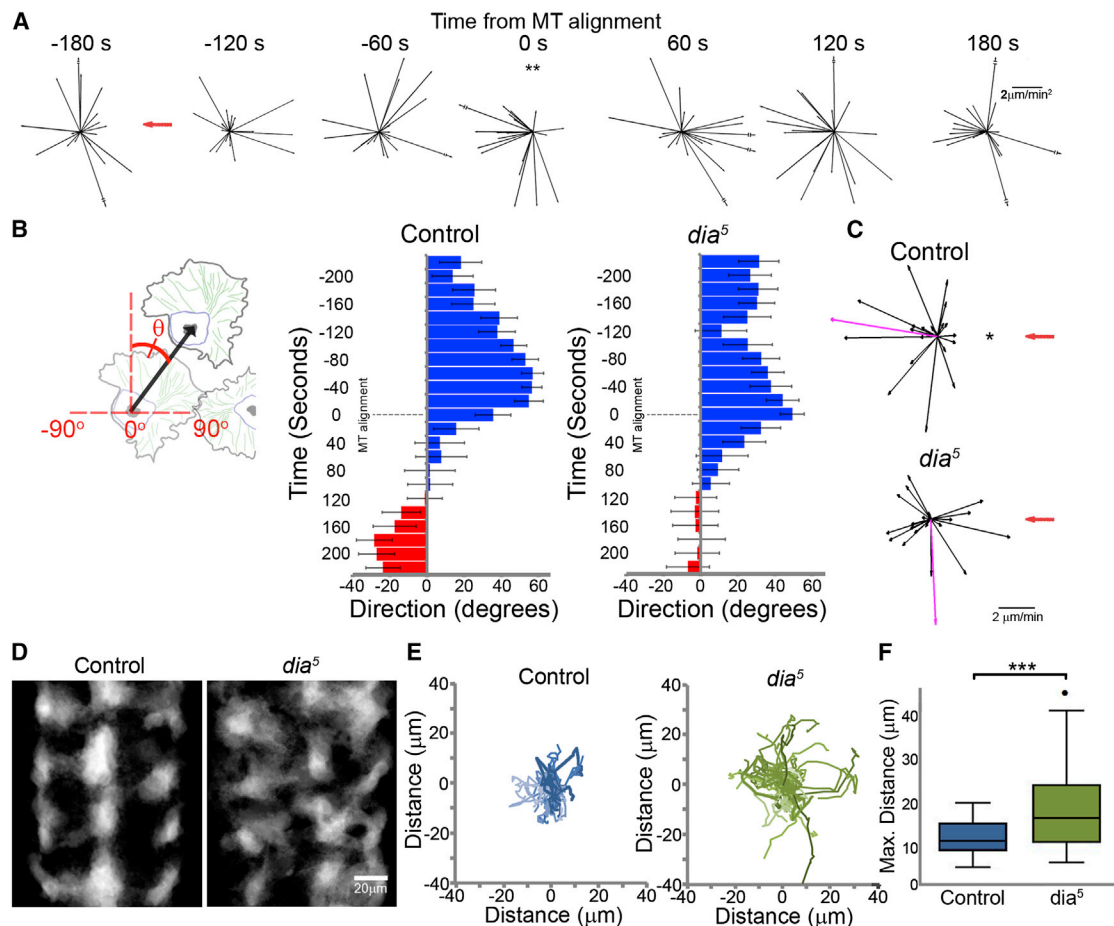
(F) Graph of mean time of lamellae contact revealed that *zip*<sup>1</sup> and *dia*<sup>5</sup> mutants maintained cell-cell contacts for a longer duration than control cells. Error bars represent S.D. \*\*p < 0.01.

(G–I) *diaphanous* mutant (*dia*<sup>5</sup>) collisions analyzed as in (A), (B), and (C).

See also Figures S4, S5, and S6 and Movie S6.

intervals (Figure S7A) suggesting their response was not as tightly coordinated as controls (Figure S1A). Furthermore, after microtubule alignment the *dia*<sup>5</sup> mutant cells showed no obvious movement away from their colliding partners (Figures 7B and S7B). We also quantified the velocities 240 s after microtubule alignment, which corresponds to the time when both control and *dia*<sup>5</sup> mutant cells have separated (Figure 6F). This revealed that while control cells migrated away from the collision, *dia*<sup>5</sup> mutants showed no significant directional preference with respect to their colliding partners (Figure 7C). These data suggest that *dia*<sup>5</sup> mutant hemocytes fail to actively repel each other.

We next determined how the alteration in cell repulsion in *dia*<sup>5</sup> mutant hemocytes affected their ability to form an evenly spaced pattern during their developmental dispersal. Analysis of the average regions occupied by hemocytes during their dispersal revealed that control cells migrated within defined domains along the ventral surface of the embryo (Figure 7D). In contrast, *dia*<sup>5</sup> mutant hemocytes showed an aberrant domain pattern (Figure 7D) and cells patrolled for greater distances within the embryo (Figures 7E and 7F), suggesting that they were less confined in their motility. *dia*<sup>5</sup> mutant hemocytes also had a slight reduction in cell spacing as shown by their decreased nearest neighbor distances (median = 17.5  $\mu\text{m}$  for controls and 15.8  $\mu\text{m}$  for *dia*<sup>5</sup>, p < 0.05) (Figures S7C and S7D; Movie S7). This suggests that precise repulsion dynamics are helping confine the migration of cells to defined regions within the embryo. Previous mathematical modeling of hemocyte dispersal suggested that tightly controlled CIL behavior was



**Figure 7. A Coordinated CIL Response Is Necessary Hemocyte Patterning**

(A) Time course of hemocyte accelerations in *dia*<sup>5</sup> mutants (black arrows) surrounding a collision event with reference to the colliding partner (red arrow). All time points show random accelerations except the time of microtubule alignment. \*\*p < 0.01.

(B) Quantification of average cell direction during the CIL time course as highlighted in the schematic. Blue highlights forward movement and red movement away from the colliding partner. Error bars represent SD.

(C) Cell velocities at 240 s after microtubule alignment with respect to the colliding partner (red arrow). Magenta arrows are the resultant velocities. Note that only controls show a significant movement away from the colliding partner. \*p < 0.05.

(D) The average regions occupied by hemocytes during their developmental dispersal revealed a disruption in the even spacing in *diaphanous* mutants.

(E) Tracks of hemocytes migrating over a 20-min period after they have spread throughout the embryo.

(F) Quantification of the maximum distance hemocytes migrate from the tracks measured in (E) revealed that *dia*<sup>5</sup> mutants migrate over greater distances in the embryo. \*\*\*p < 0.001.

See also Figure S7 and Movie S7.

essential for their normal dispersal (Davis et al., 2012). As hemocytes in *dia*<sup>5</sup> mutants showed no directional preference with respect to their colliding partners upon cell separation (Figure 7C), we wanted to determine how this reflected on the overall cell distribution in the simulation. Indeed, randomizing the sensitivity of simulated cells to the direction of their colliding partners (which in control simulations is a fixed parameter) led to a similar acquisition of aberrant domains (Figures S7E and S7F; Movie S7) and a slight reduction in cell spacing (median = 24.2  $\mu\text{m}$  for wild-type parameters and 23.3  $\mu\text{m}$  for randomized repulsion, p < 0.001). These data show that a precisely orchestrated CIL response in hemocytes is essential for it to behave as an efficient patterning cue.

## DISCUSSION

Here, we show that hemocyte dispersal requires a precisely orchestrated CIL response; cells are not stochastically repelling each other. Hemocyte collision and subsequent repulsion involves a stereotyped sequence of kinematic stages that are modulated by synchronized changes in actin and microtubule dynamics. These integrated cytoskeletal changes are regulated by a transient inter-cellular adhesion, which physically couples the actin networks of colliding cells and builds up lamellar tension. It is this inter-cellular actin-clutch and the mechanical response to the collision that allows for a precise orchestration of cellular behaviors during CIL.

In recent years, there has been speculation that clutch-like mechanisms also exist at cell-cell junctions (Giannone et al., 2009). Indeed, engagement of cell-cell adhesion molecules in neuronal growth cones leads to similar slowing of actin retrograde flow (Schaefer et al., 2008). Furthermore, apical constriction during gastrulation, which is driven by acto-myosin contraction, is hypothesized to be induced by a “clutch-like” adhesion (Roh-Johnson et al., 2012). During CIL in hemocytes, this cell-cell adhesion is critical to orchestrate both intracellular responses and inter-cellular behaviors, which allows the response to be coordinated in colliding cells. As apical constriction during gastrulation is also coordinated in space and time across numerous cells within an epithelial sheet (Martin et al., 2010), it is possible that a similar clutch-like mechanism is allowing such inter-cellular integration of forces.

The engagement of a transient inter-cellular adhesion is characteristic of CIL in a number of cell-types (Gloushankova et al., 1998; Heaysman and Pegrum, 1973; Theveneau et al., 2010). This transient adhesion is very similar to the initial punctate adhesion between epithelial cells prior to their formation of a mature cell-cell junction, which also involves a radial actin bundle running perpendicular to the leading edge (Adams et al., 1998; Gloushankova et al., 1998). These transient cadherin-based adhesions, which have been called focal adherens junctions (Huveneers and de Rooij, 2013), depend on the development of tension (Huveneers et al., 2012). Indeed, cadherins in epithelial cells, astrocytes, and fibroblasts are observed to flow in a retrograde fashion with the actin network (Peglion et al., 2014), which has led to speculation that cadherin-based cell-cell adhesions could lead to an analogous clutch-like mechanism during maturation of adherens junctions (Giannone et al., 2009). It will be interesting to investigate why epithelial cells transform a focal adherens junction into a stable cell adhesion, whereas in fibroblasts and hemocytes the adhesion results in active repulsion.

Subsequent to adhesion engagement during CIL, we observe a sudden and synchronous reorganization of the colliding actin networks. The engagement of this actin-clutch develops tension between colliding cells, which we hypothesize causes the sudden forward acceleration that we observed immediately upon lamellar overlap. It is intriguing to note that chick heart fibroblasts similarly have a momentary acceleration toward each other during their CIL response (Abercrombie and Heaysman, 1953). This intracellular tension may also help form the transient stress fiber that couples the colliding cells, as stress fiber formation is a tension responsive process (BurrIDGE and Wittchen, 2013). This creates a mechanism of haptic feedback whereby the cells “pull” on each other with the stress fiber acting as a mechanosensor during collisions, similar to its hypothesized role in sensing substrate stiffness (Trichet et al., 2012). The contraction of this actin fiber, which must be embedded within the lamellar actin network, would also explain the network-wide reorganization of actin flows in a process analogous to the “contractile treadmill” observed toward regions of actomyosin contraction in fibroblasts (Rossier et al., 2010).

As the actin networks reorganize in colliding hemocytes, microtubules polymerize into the region of low retrograde flow (that also correlates with the region of stress fiber development). Microtubules polymerize toward the leading edge in a number of

cell-types and undergo frequent catastrophes as they fight against the flowing actin network (Waterman-Storer and Salmon, 1997). In growth cones, upon adhesion engagement, the actin retrograde flow is slowed, allowing microtubules to polymerize toward the contact site (Schaefer et al., 2008) analogous to the initial stages of hemocyte CIL. We therefore speculate that the microtubule alignment observed between colliding cells during CIL is the result of microtubules following a path of least resistance within the actin network. However, it is also possible that microtubules are coupled to the stress fiber through an actin-microtubule crosslinker. Nevertheless, there must be tight coordination of both actin and microtubules during the CIL process.

While it is clear that microtubules are critical for CIL in a number of cell-types (Kadir et al., 2011; Stramer et al., 2010), their exact role during the response is currently unknown. The synchronous back acceleration in colliding partners is strongly correlated with the time when microtubule alignment occurs, suggesting that microtubule bundles are playing a role in stopping forward movement. The microtubules also appear to be critical for the generation of the precise internuclear spacing of the cells during CIL—that is critical for the emergence of the even dispersal pattern (Davis et al., 2012). Part of this spacing is governed by the regular size of hemocyte cell bodies; we hypothesize that the remainder of the nuclear spacing is controlled by the physical and dynamic properties of the microtubules themselves. As microtubules are rigid and capable of bearing compressive loads, the internuclear spacing may be controlled by a combination of tensional elements (i.e., the acto-myosin network) and elements that resist compression (i.e., microtubules). It may be the sum of these mechanical components—in a process analogous to cellular tensegrity (Ingber, 2003)—that allows the cells to precisely define their separation distance. Another possible role for microtubules may be in regulating the cell-cell adhesion. Microtubules target cell-matrix adhesions to promote their dissolution (Kaverina et al., 1999), and an analogous process may be occurring at the cell-cell adhesion during CIL.

Finally, the retraction phase of CIL involves a seemingly synchronous lamellar response in colliding partners. It is possible that the tension generated by engagement of the actin-clutch is suddenly released due to a breaking of the cell-cell adhesion. Alternatively, the tension may cause cytoskeletal components, such as the microtubule bundle or the actin stress fiber, to undergo a sudden catastrophe. Either way, it is this sudden release of lamellar tension that causes the synchronous retraction of colliding partners, which is essential to generate a choreographed CIL response. Lamellar retraction occurs prior to the generation of new protrusions away from the contact site, and we hypothesize that it is this sudden contraction that drives the subsequent repolarization phase of the response. Indeed, acto-myosin contractility initiates symmetry breaking and polarization in a number of cell-types (Cramer, 2010; Yam et al., 2007). It is also interesting to note that during initiation of polarized cell motility there is a propagation of actin network changes from the lamellar rear to the front (Yam et al., 2007), which we also observed during CIL initiation, although the mechanics behind this redistribution are currently unclear.

Here, we reveal that hemocyte CIL involves distinct stages, leading to both cells retracting from one another and subsequently repolarizing. It should be made clear that not all contact inhibitory cell-types undergo a similar sequence of events as CIL is not a homogenous response (Stramer et al., 2013). We favor the idea of broadly separating CIL into two types, first envisaged by Abercrombie (1970) and Vesely and Weiss (1973): type 1, which involves active retraction (e.g., hemocyte and fibroblast CIL), and type 2 in which forward movement is randomly deflected or stopped altogether (e.g., *dia*<sup>5</sup> mutant hemocytes, epithelial wound closure). While inhibition of cellular protrusions is sometimes thought to be the hallmark of CIL (Mayor and Carmona-Fontaine, 2010), Abercrombie (1970) believed that the predominant response in colliding cells is “a spasm of contraction” that “obliterates the process of ruffling.” It is also interesting to note the initial description of CIL by Abercrombie and Heaysman (1953): “As a result of the adhesion.... they [cells] push or pull against each other to some extent; some of the energy which would normally go into movement is thereby dissipated or becomes potential energy of elastic tension between the cells. When an adhesion breaks, the release of potential energy stored as elastic tension produces the sudden acceleration.” An inter-cellular actin-clutch is an ideal candidate to be responsible for such a response.

## EXPERIMENTAL PROCEDURES

### Microscopy

Embryos were mounted as previously described (Davis et al., 2012) and time-lapse images acquired every 5 s (for retrograde flow analysis) or 10 s (for kinematic and co-localization analyses) with a PerkinElmer Ultraview spinning disk microscope during developmental dispersal (stages 14–16). See [Extended Experimental Procedures](#) for a list of fly lines and a more detailed description of microscopy.

### Kinematics Analysis and Modeling

For kinematic analysis, hemocytes were labeled with nuclear and microtubule markers and time-lapse movies were acquired at 10 s/frame. Nuclei were automatically tracked using Volocity software (PerkinElmer). Microtubule alignment was used as a marker for a CIL event, and cells that had not collided with another cell for 4 min before and after the microtubule alignment were included in the analysis. The velocity and acceleration of cells was calculated as previously described (Dunn and Paddock, 1982). See [Extended Experimental Procedures](#) for a more detailed description of kinematics.

### Retrograde Flow Analysis

Time-lapse images of freely moving or colliding hemocytes containing actin labeled with either Lifeact-GFP or Moesin-cherry were acquired at 5 s/frame (when imaging actin alone) or at 10 s/frame (when imaging actin with other fluorescent probes). For collision analysis, cells were chosen such that the cells collided once over the duration of the time course. Cells were manually segmented prior to analysis. To quantify retrograde flow rates in lamellae, the cell body was manually segmented and data points within this region discarded. Pseudo-speckle analysis was performed as described previously (Betz et al., 2009). See [Extended Experimental Procedures](#) for a more detailed description of retrograde flow analysis.

### Laser Abscission

Hemocytes were labeled with UAS-LifeAct-GFP and UAS-RedStinger and imaged on an inverted 780 Zeiss LSM multi-photon confocal with a time interval of 1 s. Cells were imaged for 5–10 s and then ablated with a two-photon laser tuned to 730 nm and focused in a 0.4  $\mu\text{m}$   $\times$  1.5  $\mu\text{m}$  region, either at the edge or within the actin network for freely moving cells or at the region

of lamella overlap along the actin fiber for colliding cells. Hemocytes were then imaged for 60 s with only 1.1 s elapsing between frames surrounding the ablation. For mock ablation of colliding cells, the same protocol was performed as mentioned except the laser was switched off. See [Extended Experimental Procedures](#) for a more detailed description of modeling lamellar forces.

### Modeling Cytoskeletal Stresses

To compute the stresses inside the actin cytoskeleton in both freely moving and colliding hemocytes, the actin network was assumed to behave as a linear viscoelastic material. This work used the same model as Betz et al. (2011) to calculate the cytoskeletal forces developed by growth cones in vitro. See [Extended Experimental Procedures](#) for a more detailed description of modeling cytoskeletal stresses.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.015>.

## AUTHOR CONTRIBUTIONS

J.R.D., A.L., G.A.D., T.B., M.M., and B.M.S. contributed to the study concept and design, the analysis and interpretation of data, and the writing and revision of the manuscript. F.M., J.T., J.A.S., and Y.M. contributed to the acquisition and interpretation of data and review of the manuscript.

## ACKNOWLEDGMENTS

We would like to thank M. Peifer and D. Kiehart for fly lines. We appreciate comments on the manuscript from B. Baum, G. Charras, N. Davis, M. Dionne, P. Martin, E. Paluch, and W. Razzell. We also would like to thank N. Tapon, and P. Jordan and D. Alibhai from LRI Light Microscopy for help with laser ablations. B.M.S. was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) project grant BB/F020635/1. T.B. was supported by French Agence Nationale de la Recherche (ANR) grant ANR-11-JSV5-0002. J.R.D. was funded by the Medical Research Council (MRC). A.L. was funded by the BBSRC. We also thank the CMCBI at KCL for their support.

Received: March 14, 2014  
Revised: November 24, 2014  
Accepted: January 27, 2015  
Published: March 19, 2015

## REFERENCES

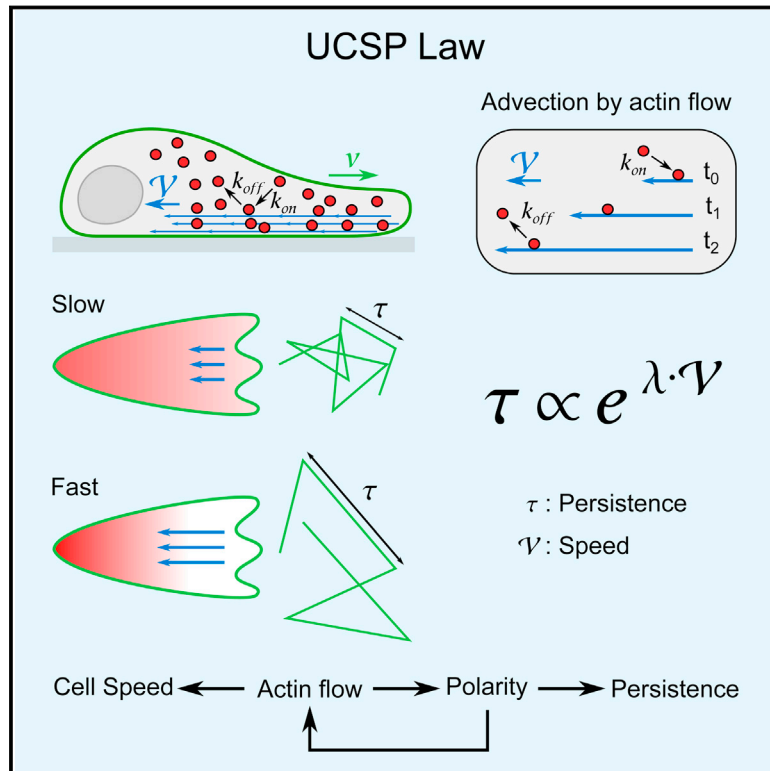
- Abercrombie, M. (1970). Control mechanisms in cancer. *Eur. J. Cancer* 6, 7–13.
- Abercrombie, M., and Heaysman, J.E. (1953). Observations on the social behaviour of cells in tissue culture. I. Speed of movement of chick heart fibroblasts in relation to their mutual contacts. *Exp. Cell Res.* 5, 111–131.
- Adams, C.L., Chen, Y.T., Smith, S.J., and Nelson, W.J. (1998). Mechanisms of epithelial cell-cell adhesion and cell compaction revealed by high-resolution tracking of E-cadherin-green fluorescent protein. *J. Cell Biol.* 142, 1105–1119.
- Astin, J.W., Batson, J., Kadir, S., Charlet, J., Persad, R.A., Gillatt, D., Oxley, J.D., and Nobes, C.D. (2010). Competition amongst Eph receptors regulates contact inhibition of locomotion and invasiveness in prostate cancer cells. *Nat. Cell Biol.* 12, 1194–1204.
- Betz, T., Koch, D., Lim, D., and Käs, J.A. (2009). Stochastic actin polymerization and steady retrograde flow determine growth cone advancement. *Biophys. J.* 96, 5130–5138.
- Betz, T., Koch, D., Lu, Y.B., Franze, K., and Käs, J.A. (2011). Growth cones as soft and weak force generators. *Proc. Natl. Acad. Sci. USA* 108, 13420–13425.



- Burridge, K., and Wittchen, E.S. (2013). The tension mounts: stress fibers as force-generating mechanotransducers. *J. Cell Biol.* 200, 9–19.
- Carmona-Fontaine, C., Matthews, H.K., Kuriyama, S., Moreno, M., Dunn, G.A., Parsons, M., Stern, C.D., and Mayor, R. (2008). Contact inhibition of locomotion in vivo controls neural crest directional migration. *Nature* 456, 957–961.
- Cramer, L.P. (2010). Forming the cell rear first: breaking cell symmetry to trigger directed cell migration. *Nat. Cell Biol.* 12, 628–632.
- Davis, J.R., Huang, C.Y., Zanet, J., Harrison, S., Rosten, E., Cox, S., Soong, D.Y., Dunn, G.A., and Stramer, B.M. (2012). Emergence of embryonic pattern through contact inhibition of locomotion. *Development* 139, 4555–4560.
- Doyle, A.D., Kutys, M.L., Conti, M.A., Matsumoto, K., Adelstein, R.S., and Yamada, K.M. (2012). Micro-environmental control of cell migration—myosin IIA is required for efficient migration in fibrillar environments through control of cell adhesion dynamics. *J. Cell Sci.* 125, 2244–2256.
- Dunn, G.A., and Paddock, S.W. (1982). Analysing the motile behaviour of cells: a general approach with special reference to pairs of cells in collision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 299, 147–157.
- Gardel, M.L., Shin, J.H., MacKintosh, F.C., Mahadevan, L., Matsudaira, P., and Weitz, D.A. (2004). Elastic behavior of cross-linked and bundled actin networks. *Science* 304, 1301–1305.
- Gardel, M.L., Sabass, B., Ji, L., Danuser, G., Schwarz, U.S., and Waterman, C.M. (2008). Traction stress in focal adhesions correlates biphasically with actin retrograde flow speed. *J. Cell Biol.* 183, 999–1005.
- Giannone, G., Dubin-Thaler, B.J., Rossier, O., Cai, Y., Chaga, O., Jiang, G., Beaver, W., Döbereiner, H.G., Freund, Y., Borisy, G., and Sheetz, M.P. (2007). Lamellipodial actin mechanically links myosin activity with adhesion-site formation. *Cell* 128, 561–575.
- Giannone, G., Mège, R.M., and Thoumine, O. (2009). Multi-level molecular clutches in motile cell processes. *Trends Cell Biol.* 19, 475–486.
- Gloushankova, N.A., Krendel, M.F., Alieva, N.O., Bonder, E.M., Feder, H.H., Vasiliev, J.M., and Gelfand, I.M. (1998). Dynamics of contacts between lamellae of fibroblasts: essential role of the actin cytoskeleton. *Proc. Natl. Acad. Sci. USA* 95, 4362–4367.
- Heaysman, J.E., and Pegrum, S.M. (1973). Early contacts between fibroblasts. An ultrastructural study. *Exp. Cell Res.* 78, 71–78.
- Hirata, H., Tatsumi, H., and Sokabe, M. (2008). Zyxin emerges as a key player in the mechanotransduction at cell adhesive structures. *Commun. Integr. Biol.* 1, 192–195.
- Homem, C.C., and Peifer, M. (2009). Exploring the roles of diaphanous and enabled activity in shaping the balance between filopodia and lamellipodia. *Mol. Biol. Cell* 20, 5138–5155.
- Hotulainen, P., and Lappalainen, P. (2006). Stress fibers are generated by two distinct actin assembly mechanisms in motile cells. *J. Cell Biol.* 173, 383–394.
- Huveneers, S., and de Rooij, J. (2013). Mechanosensitive systems at the cadherin-F-actin interface. *J. Cell Sci.* 126, 403–413.
- Huveneers, S., Oldenburg, J., Spanjaard, E., van der Krogt, G., Grigoriev, I., Akhmanova, A., Rehmann, H., and de Rooij, J. (2012). Vinculin associates with endothelial VE-cadherin junctions to control force-dependent remodeling. *J. Cell Biol.* 196, 641–652.
- Ingber, D.E. (2003). Tensegrity I. Cell structure and hierarchical systems biology. *J. Cell Sci.* 116, 1157–1173.
- Ji, L., Lim, J., and Danuser, G. (2008). Fluctuations of intracellular forces during cell protrusion. *Nat. Cell Biol.* 10, 1393–1400.
- Kadir, S., Astin, J.W., Tahtamouni, L., Martin, P., and Nobes, C.D. (2011). Microtubule remodelling is required for the front-rear polarity switch during contact inhibition of locomotion. *J. Cell Sci.* 124, 2642–2653.
- Kaverina, I., Krylyshkina, O., and Small, J.V. (1999). Microtubule targeting of substrate contacts promotes their relaxation and dissociation. *J. Cell Biol.* 146, 1033–1044.
- Martin, A.C., Gelbart, M., Fernandez-Gonzalez, R., Kaschube, M., and Wieschaus, E.F. (2010). Integration of contractile forces during tissue invagination. *J. Cell Biol.* 188, 735–749.
- Mayor, R., and Carmona-Fontaine, C. (2010). Keeping in touch with contact inhibition of locomotion. *Trends Cell Biol.* 20, 319–328.
- Nakano, K., Takaishi, K., Kodama, A., Mammoto, A., Shiozaki, H., Monden, M., and Takai, Y. (1999). Distinct actions and cooperative roles of ROCK and mDia in Rho small G protein-induced reorganization of the actin cytoskeleton in Madin-Darby canine kidney cells. *Mol. Biol. Cell* 10, 2481–2491.
- Peglion, F., Llense, F., and Etienne-Manneville, S. (2014). Adherens junction treadmill during collective migration. *Nat. Cell Biol.* 16, 639–651.
- Roh-Johnson, M., Shemer, G., Higgins, C.D., McClellan, J.H., Werts, A.D., Tulu, U.S., Gao, L., Betzig, E., Kiehart, D.P., and Goldstein, B. (2012). Triggering a cell shape change by exploiting preexisting actomyosin contractions. *Science* 335, 1232–1235.
- Rossier, O.M., Gauthier, N., Biais, N., Vonnegut, W., Fardin, M.A., Avigan, P., Heller, E.R., Mathur, A., Ghassemi, S., Koeckert, M.S., et al. (2010). Force generated by actomyosin contraction builds bridges between adhesive contacts. *EMBO J.* 29, 1055–1068.
- Sandbo, N., Ngam, C., Torr, E., Kregel, S., Kach, J., and Dulin, N. (2013). Control of myofibroblast differentiation by microtubule dynamics through a regulated localization of mDia2. *J. Biol. Chem.* 288, 15466–15473.
- Schaefer, A.W., Schoonderwoert, V.T., Ji, L., Medeiros, N., Danuser, G., and Forscher, P. (2008). Coordination of actin filament and microtubule dynamics during neurite outgrowth. *Dev. Cell* 15, 146–162.
- Stramer, B., Moreira, S., Millard, T., Evans, I., Huang, C.Y., Sabet, O., Milner, M., Dunn, G., Martin, P., and Wood, W. (2010). Clasp-mediated microtubule bundling regulates persistent motility and contact repulsion in *Drosophila* macrophages in vivo. *J. Cell Biol.* 189, 681–689.
- Stramer, B.M., Dunn, G.A., Davis, J.R., and Mayor, R. (2013). Rediscovering contact inhibition in the embryo. *J. Microsc.* 251, 206–211.
- Svitkina, T.M., Verkhovsky, A.B., McQuade, K.M., and Borisy, G.G. (1997). Analysis of the actin-myosin II system in fish epidermal keratocytes: mechanism of cell body translocation. *J. Cell Biol.* 139, 397–415.
- Theveneau, E., Marchant, L., Kuriyama, S., Gull, M., Moepps, B., Parsons, M., and Mayor, R. (2010). Collective chemotaxis requires contact-dependent cell polarity. *Dev. Cell* 19, 39–53.
- Trichet, L., Le Digabel, J., Hawkins, R.J., Vedula, S.R., Gupta, M., Ribault, C., Hersen, P., Voituriez, R., and Ladoux, B. (2012). Evidence of a large-scale mechanosensing mechanism for cellular adaptation to substrate stiffness. *Proc. Natl. Acad. Sci. USA* 109, 6933–6938.
- Vesely, P., and Weiss, R.A. (1973). Cell locomotion and contact inhibition of normal and neoplastic rat cells. *Int. J. Cancer* 11, 64–76.
- Villar-Cerviño, V., Molano-Mazón, M., Catchpole, T., Valdeolmillos, M., Henkemeyer, M., Martínez, L.M., Borrell, V., and Marín, O. (2013). Contact repulsion controls the dispersion and final distribution of Cajal-Retzius cells. *Neuron* 77, 457–471.
- Waterman-Storer, C.M., and Salmon, E.D. (1997). Actomyosin-based retrograde flow of microtubules in the lamella of migrating epithelial cells influences microtubule dynamic instability and turnover and is associated with microtubule breakage and treadmill. *J. Cell Biol.* 139, 417–434.
- Yam, P.T., Wilson, C.A., Ji, L., Hebert, B., Barnhart, E.L., Dye, N.A., Wiseman, P.W., Danuser, G., and Theriot, J.A. (2007). Actin-myosin network reorganization breaks symmetry at the cell rear to spontaneously initiate polarized cell motility. *J. Cell Biol.* 178, 1207–1221.
- Young, P.E., Richman, A.M., Ketchum, A.S., and Kiehart, D.P. (1993). Morphogenesis in *Drosophila* requires nonmuscle myosin heavy chain function. *Genes Dev.* 7, 29–41.

# Actin Flows Mediate a Universal Coupling between Cell Speed and Cell Persistence

## Graphical Abstract



## Authors

Paolo Maiuri,  
Jean-François Rupprecht, ...,  
Michael Sixt, Raphaël Voituriez

## Correspondence

matthieu.piel@curie.fr (M.P.),  
sixt@ist.ac.at (M.S.),  
voiturie@lptmc.jussieu.fr (R.V.)

## In Brief

Despite the fact that different cell types follow distinct migration patterns, the locomotion of all cellular types follows one simple universal rule: the straightness of movement (persistence) is an exponential function of speed. This general law of cell migration is explained by a physical model based on the transport of polarity factors by the actin retrograde flows and predicts a diagram of possible cell trajectories.

## Highlights

- Speed and persistence of cells are exponentially correlated (the UCSP law)
- Faster cells turn less: speed stabilizes cell directionality
- The UCSP originates from transport of polarity factors by the retrograde actin flow
- A physical model explains the UCSP and predicts a diagram of cell trajectories



# Actin Flows Mediate a Universal Coupling between Cell Speed and Cell Persistence

Paolo Maiuri,<sup>1,9</sup> Jean-François Rupprecht,<sup>2,9</sup> Stefan Wieser,<sup>3,9</sup> Verena Ruprecht,<sup>3</sup> Olivier Bénichou,<sup>2</sup> Nicolas Carpi,<sup>1</sup> Mathieu Coppey,<sup>4</sup> Simon De Beco,<sup>4</sup> Nir Gov,<sup>5</sup> Carl-Philipp Heisenberg,<sup>3</sup> Carolina Lage Crespo,<sup>6</sup> Franziska Lautenschlaeger,<sup>1</sup> Maël Le Berre,<sup>1</sup> Ana-Maria Lennon-Dumenil,<sup>7</sup> Matthew Raab,<sup>1</sup> Hawa-Racine Thiam,<sup>1</sup> Matthieu Piel,<sup>1,\*</sup> Michael Sixt,<sup>3,\*</sup> and Raphaël Voituriez<sup>2,8,\*</sup>

<sup>1</sup>Institut Curie, CNRS UMR 144, 26 rue d'Ulm, 75005 Paris, France

<sup>2</sup>Laboratoire de Physique Théorique de la Matière Condensée, UMR 7600 CNRS /UPMC, 4 Place Jussieu, 75255 Paris Cedex, France

<sup>3</sup>Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria

<sup>4</sup>Institut Curie, CNRS UMR 168, 26 rue d'Ulm, 75005 Paris, France

<sup>5</sup>Department of Chemical Physics, Weizmann Institute of Science, 76100 Rehovot, Israel

<sup>6</sup>Division of Immunology, Transplantation and Infectious Diseases, San Raffaele Scientific Institute, 20132 Milan, Italy

<sup>7</sup>U 653, Inserm/Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France

<sup>8</sup>Laboratoire Jean Perrin, UMR 8237 CNRS /UPMC, 4 Place Jussieu, 75255 Paris Cedex, France

<sup>9</sup>Co-first author

\*Correspondence: [matthieu.piel@curie.fr](mailto:matthieu.piel@curie.fr) (M.P.), [sixt@ist.ac.at](mailto:sixt@ist.ac.at) (M.S.), [voituriez@lptmc.jussieu.fr](mailto:voituriez@lptmc.jussieu.fr) (R.V.)

<http://dx.doi.org/10.1016/j.cell.2015.01.056>

## SUMMARY

Cell movement has essential functions in development, immunity, and cancer. Various cell migration patterns have been reported, but no general rule has emerged so far. Here, we show on the basis of experimental data *in vitro* and *in vivo* that cell persistence, which quantifies the straightness of trajectories, is robustly coupled to cell migration speed. We suggest that this universal coupling constitutes a generic law of cell migration, which originates in the advection of polarity cues by an actin cytoskeleton undergoing flows at the cellular scale. Our analysis relies on a theoretical model that we validate by measuring the persistence of cells upon modulation of actin flow speeds and upon optogenetic manipulation of the binding of an actin regulator to actin filaments. Beyond the quantitative prediction of the coupling, the model yields a generic phase diagram of cellular trajectories, which recapitulates the full range of observed migration patterns.

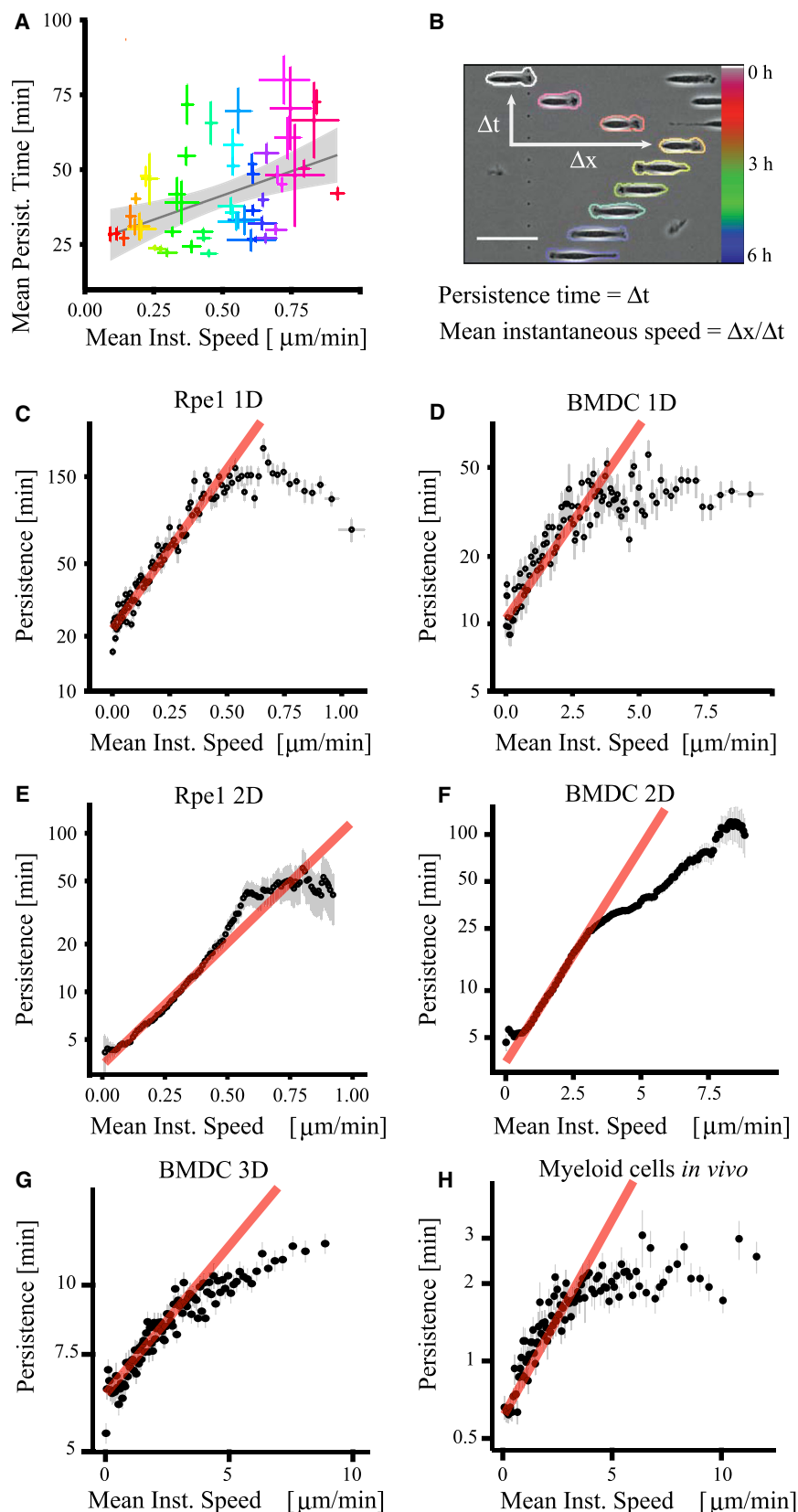
## INTRODUCTION

Eukaryotic cell migration is essential for a large set of biological processes. Assessing quantitatively the exploratory efficiency of cell trajectories is therefore crucial. In the absence of external guidance, cell movement can be described as a random motion, and proposed models have ranged from simple Brownian motion to persistent random walks (Selmeczi et al., 2008), Levy walks (Harris et al., 2012), or composite processes such as intermittent random walks (Bénichou et al., 2011). Such models differ in the cell persistence, which quantifies the ability of a cell to maintain its direction of motion. The variety of behaviors, observed even along a single cell trajectory, stems from the

fact that, as opposed to a passive tracer in a medium at thermal equilibrium, which performs a classical Brownian motion, a cell is self-propelled, and as such, belongs to the class of active Brownian particles (Romanczuk et al., 2012). This class of processes is extremely vast and needs to be restricted to have some predicting power. Up to now universal behaviors have emerged in the context of the collective dynamics of self-propelled particles (Czirók et al., 1998; Toner et al., 2005; Vedel et al., 2013), but remain elusive at the level of single cells.

Recently, a vast amount of data of individual cell trajectories has been collected over many cell types in the context of the First World Cell Race (Maiuri et al., 2012) (Figures 1A and 1B). These data show that, despite an apparent diversity at the level of the whole population analyzed, correlations between the mean linear instantaneous cell speed and the persistence time, defined as the mean time during which a cell maintains its direction of motion, exist. This general trend that faster cells migrate more straight than slower cells, which has since been observed by others (Wu et al., 2014), remains unexplained and suggests that robust mechanisms could constrain the possible characteristics of cell trajectories.

In this work, we analyzed trajectories of cells migrating in various *in vitro* assays and in live tissues. We reveal on the basis of this extensive data set a universal coupling between cell speed and cell persistence (UCSP). To explain what appears as a universal law of cell migration, we developed a physical model relying on minimal hypothesis that shows that actin flows, which are the hallmark of motile cells (Theriot and Mitchison, 1991; Svitkina et al., 1997; Wilson et al., 2010), reinforce cell polarity and consequently cell persistence. The model was validated experimentally by two assays: one enabling a gradual modulation of actin flow speeds combined with pharmacological interference experiments, and a second, based on an engineered optogenetic molecular module, which allowed controlling an actin polymerization regulator, Arpin. Finally, the model has the following merits: it (1) quantitatively predicts the observed exponential correlation between speed and



**Figure 1. Correlation between Cell Persistence and Cell Speed**

(A) Population mean persistence time versus mean instantaneous speed. Data from the First World Cell Race, refers to the original article (Maiuri et al., 2012) for the cell-type color code. Linear fit (solid line) and 0.95 confidence interval (gray).

(B) Definition of mean persistence time and mean instantaneous speed in 1D. Cell contour color shows the time progression. Scale bar, 50  $\mu\text{m}$ .

(C–F) Persistence time, binned for the corresponding instantaneous speed, versus instantaneous speed. (C) RPE1 cells on micropatterned lines of 9- $\mu\text{m}$  width coated with fibronectin. (D) BMDCs in fibronectin-treated channels with a  $7 \times 5 \mu\text{m}^2$  square section. (E–F) The persistence time in 2D is defined here as the time needed for a cell to change its original direction of motion by  $90^\circ$ . (E) Data of RPE1 cells on 2D surface uniformly treated with fibronectin. (F) BMDCs confined between two parallel, fibronectin-treated planes, 5  $\mu\text{m}$  apart from each other.

(G) BMDCs embedded in bovine collagen gel and confined between two planes 5  $\mu\text{m}$  apart from each other.

(H) Myeloid cells imaged in live Medaka fish.

Red curves represent the exponential fit of the experimental data. Black dots and gray lines are mean and SE for the binned data on both axis.

See also Figures S1 and S2 and Movie S1.



persistence that characterizes the UCSP, (2) provides from minimal microscopic hypothesis an explicit construction of single cells dynamics as active Brownian particles, and (3) yields a generic phase diagram of cellular trajectories, which recapitulates the full range of observed migration behaviors.

## RESULTS

### Cell Trajectory Analysis Reveals a Universal Coupling between Cell Speed and Persistence

The First World Cell Race (Maiuri et al., 2012), which gathered recordings of individual cell trajectories on 1D adhesive tracks for 54 different adherent cell types made available an unprecedented amount of data. We performed a further analysis of this data and confirmed a clear positive correlation between the population averaged mean linear instantaneous cell speed and the population averaged persistence time at the level of all cell types (Figure 1A), despite their variety. To assess the robustness of this observation and check its validity at the level of single cell trajectories, we performed new experiments on two representative examples of migrating cell with either a mesenchymal migration mode (hTERT-immortalized retinal pigment epithelial cell line [RPE1]) or an amoeboid migration mode (immature bone marrow-derived mice dendritic cells [BMDC]), in various geometries (Figures S1 and S2; Movie S1): 1D adhesive tracks (Figure 1C), 1D microchannels (Figure 1D), 2D non-confined adhesive substrates (Figure 1E) and 2D confined substrates (Figure 1F). Very long cell trajectories were recorded (see Supplemental Information), allowing a clear assessment of individual cell persistence. In all cases, for which cell concentrations were low enough to treat cell trajectories as independent, we found a striking correlation between persistence time  $\tau$  and mean instantaneous velocity  $v$ , which is well fitted by a simple exponential curve  $\tau = Ae^{2v}$  before eventually saturating at larger  $v$  (see Discussion). We then checked whether this correlation was still valid when cells were migrating in more complex environments. We recorded BMDCs migrating in 3D collagen gels (Figure 1G) and myeloid cells migrating in live Medaka fish (Grabher et al., 2007) (Figure 1H). In both cases, persistence and speed were still highly correlated. Overall, these results show that the UCSP is robust when tested in different migration environments and organisms, and that the correlation, observed at the population level or at the level of individual cells exists all along a cell trajectory: a faster displacement is correlated with a straighter path. The analysis of this remarkably robust correlation that defines the UCSP is at the core of this work.

### Faster Actin Retrograde Flow Lengthens Cell Persistence Time

To elucidate the origin of the UCSP, we reasoned that, as this law applies to all cell types we tested, including cells with very different modes of migration (mesenchymal and amoeboid cells), it had to rely on a very conserved aspect of cell locomotion. Even if many details might vary from one cell type to the other, the most conserved aspect of cell locomotion is the retrograde translocation of actin filaments in the frame of reference of the cell, from the front to the rear of its locomotory parts. Actin flows can either exist over large portions of the cell (in particular

in amoeboid cells) (Renkawitz et al., 2009), or in some cases be essentially limited to the protrusive parts (such as lamellipodia in mesenchymal cells) (Theriot and Mitchison, 1991; Svitkina et al., 1997; Wilson et al., 2010). This retrograde movement is powered by the combined forces of actin polymerization at the leading edge and actomyosin contraction at the trailing edge and represents the driving force for locomotion. Upon coupling to the environment either via transmembrane adhesion receptors of the integrin family or via friction, retrograde actin flow is turned into traction forces, which pull the cell forward while actin filaments slide to the back of the cell (Mogilner and Oster, 1996; Jurado et al., 2005; Hawkins et al., 2011).

To test if actin flows are involved in the coupling between cell persistence and cell speed, we used mature bone marrow dendritic cells (mBMDCs), for which it was already demonstrated that actin retrograde flow can be varied without affecting cell speed when substrate adhesion strength is modulated (Renkawitz et al., 2009). Indeed, the cell speed  $v$  and the velocity of the actin flow  $V$  (defined hereafter in the frame of reference of the moving cell) are usually linearly coupled according to  $v = \alpha V$  (Jurado et al., 2005), such linear approximation being valid at least in the lower range of velocities for each cell type. The coefficient  $\alpha$  models the effective friction between actin filaments and the substrate, usually mediated by specific adhesion proteins (Mitchison and Kirschner, 1988; Gardel et al., 2010) and therefore depends on experimental conditions. In some cell types, varying the actin/substrate coupling parameterized by  $\alpha$  induces an adaptive response, which allows a cell to keep the speed of locomotion  $v$  relatively constant, despite different retrograde velocities  $V$  of the loosely coupled actin network. Such adaptation has been extensively characterized in mBMDCs. When placed in confined environments, these cells can flexibly shift between integrin-independent and integrin-dependent force transduction as even in the absence of these adhesion receptors the cells are able to generate sufficient traction to migrate. However, depletion of integrins or their ligands and the associated drop in friction causes retrograde actin slippage, which is then compensated by up to 2-fold increase in actin polymerization speed (Renkawitz et al., 2009). This property of mBMDCs moving in confined environments allowed us to independently study the influence of retrograde actin flow and actual cell speed on migratory persistence. Surface adhesion was independently controlled by depletion of  $\beta 2$  integrins and their ligands (coating the surfaces with an inert PEG layer), while cell speed was gradually varied by temperature changes and by pharmacological inhibition of actomyosin contractility (blebbistatin). This allowed us to tune retrograde flows from 3 to 15  $\mu\text{m}/\text{min}$  independently of cell speed (Figure 2A). Strikingly, we observed that the mean persistence time  $\tau$  measured for each experimental condition was strongly correlated with the mean speed  $V$  of the actin retrograde flow and was well-fitted by a simple exponential  $\tau = A'e^{2'V}$  (Figure 2B; Movie S2), whereas no correlation between  $\tau$  and cell speed  $v$  was observed. Note, however, that within each experimental condition ( $\alpha$  fixed) the linear scaling between  $v$  and  $V$  still held, so that the correlation between cell speed and cell persistence was preserved at the level of each experimental condition. These data strongly suggest that the observed UCSP originates from a coupling between cell persistence and the actin flow.

### Faster Actin Retrograde Flow Enhances the Asymmetry of Polarity Cues and Stabilizes Cell Polarity

A simple hypothesis to explain these effects of actin flows would be that, in polarized migrating cells, actin flows reinforce cell polarity by enhancing the asymmetry of polarity cues, as was proposed for the establishment of polarity in early *Caenorhabditis elegans* embryos (Munro et al., 2004; Goehring et al., 2011) (Figures 2C and 2D). Examples of such cues, which are likely to vary with cell type, could be molecules responsible for the generation of contractile stress, such as myosin motors or their activators, actin polymerization regulators, or could alternatively be involved in the regulation of microtubule dynamics (Zhang et al., 2014). We introduce a first element of physical modeling to test this hypothesis. We consider a migrating cell and denote by  $x$  the coordinate in the reference frame of the moving cell, defined by the instantaneous direction of migration and by  $L$  the cell extension along this axis (Figure 2C). We are interested in the dynamics of the concentration of a generic polarity cue  $c(x, t)$ , which we assume here depends only on  $x$ . This cue is assumed to specify the rear part of the cell, so that  $c(x=0, t) > c(x=L, t)$  for a cell moving in the  $+$  direction. The polarity cue can either diffuse in the cytosol, with diffusion coefficient  $D$ , or, depending on its affinity for actin, be advected by the cytoskeletal flow, whose velocity along  $x$  is denoted by  $-V$  (with  $V > 0$ ) in the cell reference frame. We denote by  $k_{on}$  and  $k_{off}$  the corresponding binding and unbinding rates of the cue to the actin cytoskeleton. In the limit of fast exchange ( $k_{on}, k_{off}$  large), the dynamics of  $c(x, t)$  depends on advective transport and on diffusion as follows:

$$\partial_t c(x, t) - \partial_x [\tilde{V} c(x, t)] = \tilde{D} \partial_x^2 c(x, t) + \partial_x \zeta_c, \quad (\text{Equation 1})$$

where  $\tilde{V} = V k_{on} / (k_{on} + k_{off})$  and  $\tilde{D} = D k_{off} / (k_{on} + k_{off})$ ; here  $\zeta_c$  is a Gaussian white noise that encompasses the fluctuations of the flux of cue molecules. At steady state, assuming  $V$  constant and the conservation of the total amount of cues, the mean cue concentration profile is therefore given by:

$$\bar{c}_V(x) = C e^{-\tilde{V}x/\tilde{D}}, \quad (\text{Equation 2})$$

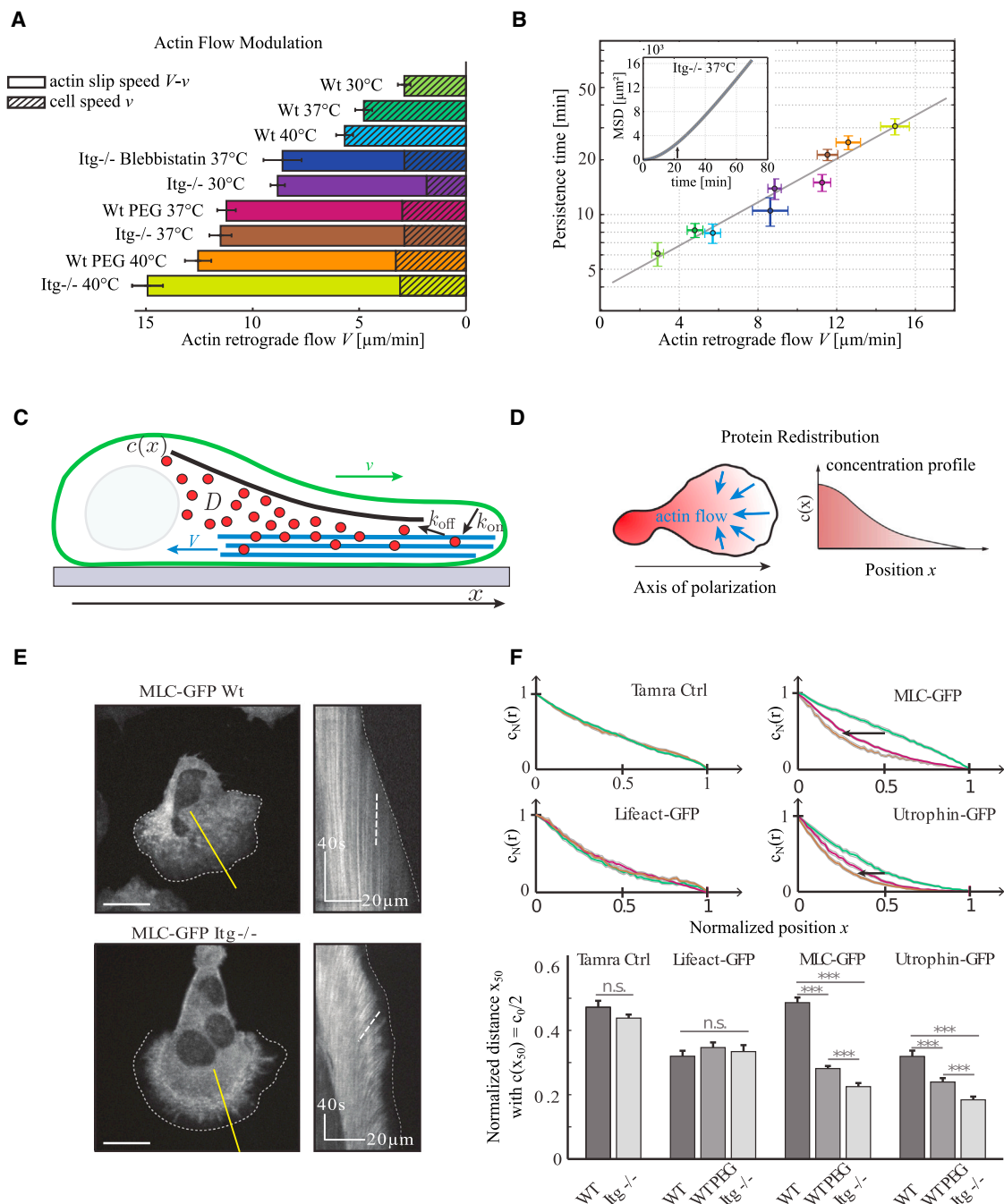
where  $C$  is a normalization constant and the dependence on the retrograde flow is denoted in subscript. This simple argument, which does not take into account heterogeneities in actin flows predicts simple exponential concentration profiles, whose steepness is controlled by the speed of the effective retrograde flow  $\tilde{V}$  (Figure 2D). Non-uniform actin flow profiles, as observed for example in (Wilson et al., 2010; Ofer et al., 2011) would change the exponential shape of concentration profiles, but would leave the dependence on the actin flow speed qualitatively unchanged. To check this general prediction, we measured on motile mBMDs the concentration profiles along the polarity axis for various molecules with different affinity for actin (Figure 2E; Movie S3): Tamra (that unspecifically labels cytoplasmic proteins), Lifeact-GFP (low affinity) (Riedl et al., 2008), MLC-GFP (high affinity) (Kondo et al., 2011), and Utrrophin-GFP (high affinity) (Burkel et al., 2007). As expected from the model, we observed that increasing the actin retrograde

flow could significantly increase the slope of the concentration profile of strong actin binders (MLC, Utrrophin), whereas the profiles of molecules with low (Lifeact) or no (Tamra) affinity to actin remained unchanged (Figures 2E and 2F). This validates our hypothesis that actin flows reinforce the asymmetry of concentration profiles of actin binding molecules.

We next reasoned that such mechanism in principle applies to any diffusing molecule that interacts with actin, and in particular, to polarity cues. It is then expected that increasing actin flows should increase the asymmetry of the concentration profile of any polarity cue, thereby stabilizing cell polarization and consequently increasing cell persistence, according to a mechanism discussed in Svitkina et al. (1997) (note that the connection to persistence was, however, absent in this reference). To further support this hypothesis, we measured the number of protrusions for each cell, using the same nine experimental conditions to vary the actin flow and found that in conditions in which cells had a faster actin flow, there was a larger proportion of cells with a single well-polarized lamellipodium (Figures 3A, 3B, and S3). We then measured the life time of such unipolar configurations, that we call polarization time  $\tau_p$ , and the actin retrograde flow speed in each protrusion, and found that they showed the same exponential correlation (Figure 3C). This suggests that the observed coupling between actin retrograde flow and cell persistence originates from the stabilization of polarity by faster actin flows, by favoring a unipolar configuration of the cell and by lengthening the protrusion lifetime.

While in principle this mechanism can apply to any diffusing polarity cue that interacts with actin, species involved in the regulation of actin flow are the most susceptible to mediate the UCSP. Cdc42, a member of the Rho GTPase family is a central regulator of polarity, mediating its function via the actin and microtubule cytoskeleton (Heasman and Ridley, 2008), and thus appeared as a natural candidate. Loss of Cdc42 in dendritic cells severely affects cell polarization and migration efficiency in vitro and in vivo (Lämmermann et al., 2009). To test if the Cdc42 polarity module was involved in coupling rearward cortical flows to polarity we selectively inhibited Cdc42 using ML-141 (Hong et al., 2013) and measured the lifetime of cell polarization  $\tau_p$  in dependence of cortical flow speed  $V$  that was modulated via a set of five independent experimental conditions as discussed above (Figure 2A). As expected, inhibition of Cdc42 led to an overall reduction of the polarization lifetime for all conditions but, importantly, longer polarization lifetimes were observed for increasing actin flow speeds  $V$  and the UCSP was preserved (Figure 3D). These results indicate that the UCSP can function independently of a Cdc42-mediated signaling pathway.

We next sought to unravel key molecular polarity cues closely affected by the cortical network flow. Myosin-II is a prominent candidate, which strongly binds to actin; it is thus transported with actin flows and tends to accumulate at the cell back when actin flows are strong (Poincloux et al., 2011; Hawkins et al., 2011). It has previously been implicated in actomyosin network organization, polarized cortical architecture and migration efficiency (Vicente-Manzanares et al., 2009). Indeed, we observed that myosin II light chain localization was strongly influenced by cortical flow modulations with a more rearward accumulation



**Figure 2. Modulation of Actin Retrograde Flow Speed Reveals a Positive Feedback Loop on the Stability of Cell Polarity**

(A) Retrograde flow speed  $V$  in the reference frame of the cell in nine different conditions. Bars represent SEM.

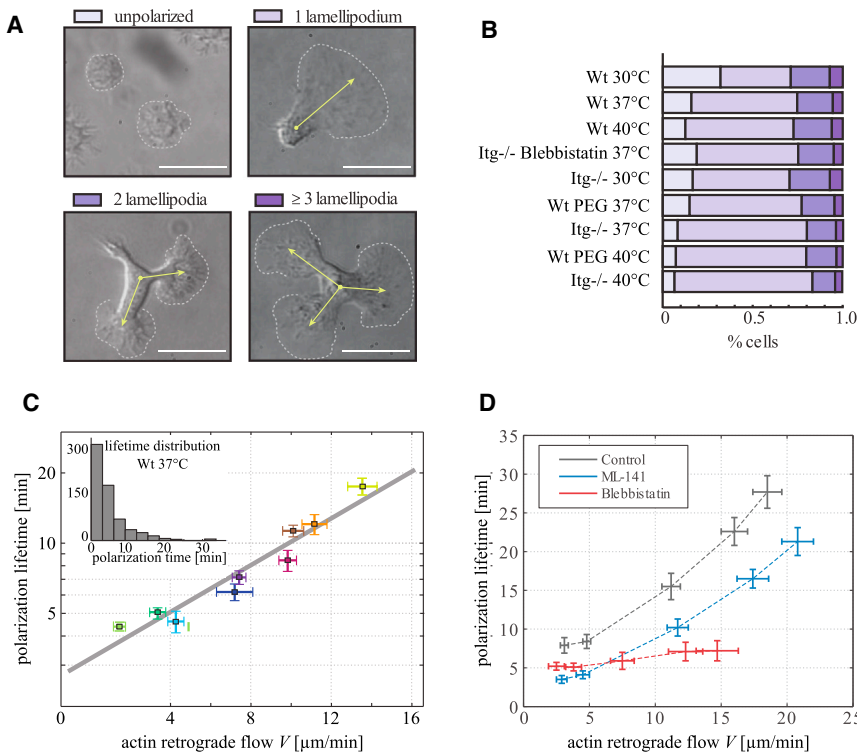
(B) Cell persistence time versus retrograde flow speed. Bars represent SEM, and the gray line represents the exponential fit. Inset shows mean-square displacement plot for  $\text{Itg}^{-/-}$  cells at  $37^\circ\text{C}$ . Arrow highlights the crossover from persistent to random motility on longer time scales.

(C) Schematic illustrating the model with a minimal set of kinetic parameters. Polarity factors are shown as red dots, actin filaments are in blue while cell outline is green and migration substrate is gray.

(D) Schematic showing the principle of protein redistribution by the actin retrograde flow. Density of transported protein is shown in red, and depends on the position  $x$  along the cell polarity axis.

(E) Left: fluorescence images of MLC-GFP localization in migrating wild-type (Wt) and  $\beta 2$  integrin knockout ( $\text{Itg}^{-/-}$ ) dendritic cells under 2D confinement. Note the enhanced depletion of MLC from the leading edge in the  $\text{Itg}^{-/-}$  cell. Dashed lines indicate cell borders. Scale bar in  $\mu\text{m}$ . Right: kymograph performed along yellow lines. White dashed lines indicate retrograde myosin flow in the lab reference frame.

(legend continued on next page)



**Figure 3. Stability of Cell Polarity as a Function of Actin Flow Speed and Effect of Pharmacological Inhibition of Myosin II and Cdc42 in mBMDCs**

(A) Differential interference contrast (DIC) representative images of cells in the unpolarized states (no lamellipodium), polarized state (one lamellipodium), and with two lamellipodia (bi-polar) and three or more lamellipodia (multi-polar). Arrows denote the axis of polarization. Scale bar, 20  $\mu$ m.

(B) Histogram of observed cell morphologies (color code see A) for nine different conditions (see Figure 2).

(C) Polarization lifetime as a function of actin flow speed. Gray line indicates exponential fits to the data. Inset shows the exponential distribution of polarization lifetimes for Wt dendritic cells at 37°C. Bars represent SEM.

(D) Polarization lifetime for control conditions (dark gray), ML-141-treated cells (Cdc42 inhibitor, blue), and Blebbistatin-treated cells (Myosin II inhibitor, red). Actin retrograde flow speed  $V$  is modulated (from left to right) using wild-type (Wt) dendritic cells (DCs) at 30°C, Wt DCs at 37°C,  $\beta$ 2 integrin knock-out ( $Itg^{-/-}$ ) DCs at 30°C,  $Itg^{-/-}$  DCs at 40°C, and  $Itg^{-/-}$  DCs at 41.5°C. Bars represent SEM. See also Figure S3 and Movies S2 and S3.

for higher actin flow speeds  $V$  (Figure 2F). To test a potential role of myosin-II in the UCSP we pharmacologically interfered with its activity by using the myosin-II inhibitor Blebbistatin. Inhibition of Myosin-II induced a reduction of cell polarization lifetimes for all conditions and polarization lifetimes  $\tau_p$  were no longer correlated to cortical flow speed  $V$  but remained at constantly low values for all  $V$  (Figure 3D). These data suggest that myosin-II constitutes an indispensable polarity cue in mBMDCs that is involved in the establishment of the UCSP.

### Physical Modeling Predicts the UCSP

To substantiate our hypothesis that the UCSP originates from the coupled dynamics of actin flow and diffusing polarity cues, we developed a minimal theoretical model, which assumes that the actin flow  $V$  is also subject to fluctuations, due for example to the stochasticity of polymerization/depolymerization processes or the heterogeneity of the environment. The model, whose main ingredients are summarized below (see Supplemental Information for details), relies on the key assumption that the mean value  $V^*$  of the actin flow (for a fixed cue concentration profile) is governed by the asymmetry of the cue concentration profile. More precisely, we assume that

$$V^* = \beta(c^*(0, t) - c^*(L, t)), \quad (\text{Equation 3})$$

where  $\beta$  is an effective parameter that controls the intensity of the coupling between the actin flow and the asymmetry of the cue concentration profile and can be interpreted as the maximal possible velocity of the actin flow. Here,  $c^*(x, t)$  denotes the fraction of activated cues, i.e., cues that induce actin flow. We show in the Supplemental Information that the phenomenological coupling (Equation 3) covers the cases where actin flows are generated by asymmetric distributions of either actin polymerization regulators (such as Arpin, see the discussion below) (Jullicher et al., 2007) or activators of contractility (such as Myosin II or a Myosin II activator, as discussed above) (Hawkins et al., 2011; Bois et al., 2011; Callan-Jones and Voituriez, 2013), which are the two main scenarios that we propose. We here do not aim at describing in details the biochemical steps involved in the process and assume a classical Hill response function of index  $n$  (results are qualitatively unchanged for other choices):

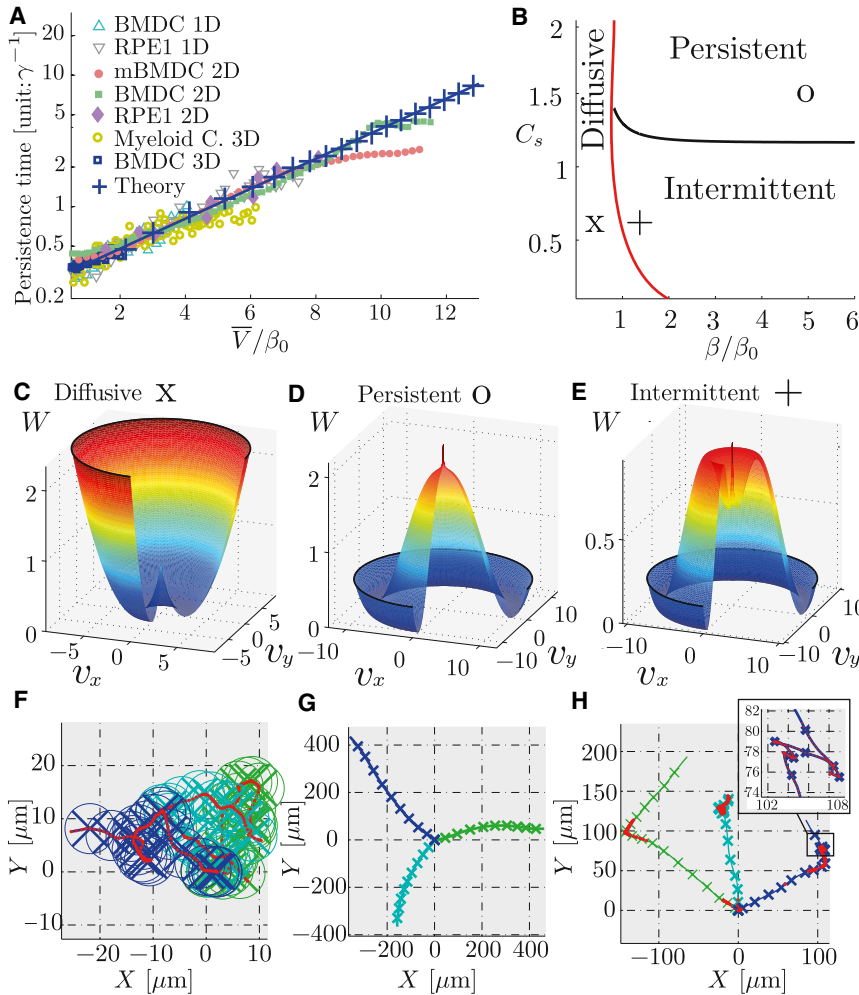
$$c^*(x, t) = \frac{c^n(x, t)}{C_s^n + c^n(x, t)}. \quad (\text{Equation 4})$$

Here,  $C_s$  is the concentration of cues above which activation is saturated and is therefore determined by the maximal concentration of activated cues.

(F) Normalized concentration profiles as a function of the normalized position  $x$  along the cell polarity axis (cell front is set at  $x = 1$ ) for a set of actin-binding molecules with various actin affinity. Concentration profiles are color coded for different cell-substrate adhesion strengths (green, Wt; magenta, Wt on Peg; brown,  $Itg^{-/-}$ , all 37°C). Arrows indicate a shift of the polarity cue concentration profile toward the cell rear with increasing actin retrograde flow. (right) Normalized distance  $x_{50}$  at which the concentration drops to  $c_{50} = c_0/2$  (\*\*\*)  $p < 0.001$ , n.s. not significant.

See also Movies S2 and S3.





**Figure 4. Model Predictions: The UCSP Law and Phase Diagram of Cell Trajectories**

The values of the parameters used in the figure (except for  $\beta, C_s$ ) and the fitting procedure are described in the [Supplemental Information](#).

(A) Persistence time as a function of the normalized mean velocity  $\bar{V}/\beta_0$ , where  $\beta_0 \equiv \bar{D}/L$  for the prediction of the model and for all available experimental data (rescaled).

(B) Phase diagram of cell trajectories in the  $\beta$  (maximal actin flow speed) and  $C_s$  (maximal concentration of activated cues) plane. Symbols correspond to the parameters used in (C) and (F) (x), (D) and (G) (O), and (E) and (H) (+).

(C–E) Effective potential  $W$  as a function of the velocities  $V_x$  and  $V_y$  normalized by  $\beta_0$ . The singularity in  $W$  at the origin  $V = 0$  is due to the anisotropy of the diffusion process ([Romanczuk et al., 2012](#)). (C)  $W$  is locally quadratic for  $V$  small: diffusive phase (marked by x in B). (D)  $W$  has a “sombbrero” shape: persistent phase (marked by O in B). (E)  $W$  has a mixed shape: intermittent phase (marked by + in B).

(F–H) Examples of simulated trajectories in the diffusive (F), persistent (G), and intermittent phases (H). Colored crosses indicate the positions at regular time intervals. Circles in (F) indicate the cell size  $L$ .

See also [Figures S4 and S5](#).

Next, we make use of the fact that the typical diffusion time of cues over the cell length  $L^2/D$  (of the order of seconds) is significantly shorter than the characteristic timescale of fluctuations in the actin flow  $\gamma^{-1}$  (of the order of minutes). The concentration of cues is then taken at steady state, and Equation 1 implies that it is a Poisson random variable, i.e.,  $c(x, t) = \bar{c}_V(x) + \delta c$  with  $\overline{\delta c^2} = K_c \bar{c}_V(x)$ , where  $K_c$  is a constant that controls the intensity of the particle number fluctuations and  $\bar{c}_V(x)$  is the steady-state profile defined in Equation 2. Considering a cell moving on a 2D substrate (the 1D case is then deduced by taking a vanishing angular diffusion, and the generalization to 3D migration is straightforward, see [Supplemental Information](#)), the dynamics of the actin flow velocity  $\mathbf{V}$ , which is a vector of modulus  $V$  and polar angle  $\phi$ , can then be written after linearization with respect to  $\delta c$  (see [Supplemental Information](#) for details):

$$\begin{aligned} \partial_t V &= \gamma \mathcal{F}(V) + \sigma(V) \zeta_V \\ \partial_t \phi &= \frac{\sqrt{K}}{V} \zeta_\phi \end{aligned} \quad (\text{Equation 5})$$

where  $\zeta_V$  and  $\zeta_\phi$  are Gaussian white noises of variance unity. Here,  $\gamma^{-1}$  is the typical timescale of the actin flow fluctuations

and  $K$  controls their amplitude, the effective force  $\mathcal{F}(V)$  and noise intensity  $\sigma(V)$  are given explicitly in the [Supplemental Information](#). The dynamics of  $V$ , fully defined by Equation 5, is therefore the dynamics of a Brownian particle in a force field  $\mathcal{F}(V)$  in the presence of a non-trivial noise with additive and multiplicative

parts. Assuming that, for any given experimental condition, the cell velocity is directly proportional to the actin flow velocity ( $\mathbf{v} = \alpha \mathbf{V}$ , where we set hereafter  $\alpha = -1$  for the sake of simplicity), Equation 5 makes it possible to fully characterize the resulting cell trajectories. In particular, the steady-state distribution  $P(v)$  of the velocity can be obtained as (see [Supplemental Information](#)):

$$P(v) = \frac{N}{\sigma^2(v)} \exp \left( 2\gamma \int_0^v du \frac{\mathcal{F}(u)}{\sigma^2(u)} \right) \equiv N e^{-\gamma W(v)}, \quad (\text{Equation 6})$$

where  $W(v)$  is an effective potential ([Figures 4C–4E](#)) and  $N$  a normalization constant, and the persistence time can be deduced from the analysis of the autocorrelation function. In turn, the polarization time  $\tau_p$  (defined as the mean lifetime of a cellular configuration, [Figure 3C](#)) can be analytically obtained as the mean first-passage time ([Condamin et al., 2007](#)) at  $v = 0$  (see [Supplemental Information](#)).

The analysis of the model reveals that the polarization time  $\tau_p(v)$  (obtained analytically) and persistence time  $\tau(v)$  (obtained from Monte Carlo simulations of Equation 5) can be very well

approximated by a simple exponential  $\tau, \tau_p \approx Ae^{\lambda v}$  in dimensions 1, 2, and 3 for a wide range of parameters as soon as the Hill index satisfies  $0.7 < n < 1.5$ , which shows that no strong nonlinearity needs to be invoked to fit the data. Remarkably, this allowed us to fit all the available data (Figures 4A and S4; Table S1) with a single universal exponential master curve and therefore reproduce quantitatively the UCSP law. This provides a first clear validation of the model.

### Synthetic Engineering of a UCSP Module Using an Optogenetic Approach

The mechanism underlying our model is that the advection of a rear-associated polarity cue from the cell leading edge to the rear (of the cell or of a protrusion) by the actin retrograde flow would contribute to the concentration of the cue at the rear and depletion from the front. This would reinforce cell polarity (or increase the lifetime of the protrusion), thus increasing cell speed and persistence, generating a positive feedback loop. To prove it experimentally, we developed a synthetic system based on Arpin and on the CRY2/CIBN optogenetic tool (Figure 5A). Arpin is a negative regulator of the Arp2/3 complex that is known to decrease lamellipodial protrusion and that does not directly bind to actin (Dang et al., 2013). CRY2 and CIBN are two independent proteins that rapidly form a heterodimer after blue light illumination (Kennedy et al., 2010). To directly test the central hypothesis of our model, we fused Arpin to CRY2 (and to mCherry to track Arpin localization), while two actin binding proteins that strongly differ in their affinity to actin, LifeAct and Utrophin-CH (Utr) (Figure 2F), were fused to CIBN (and to GFP). As expected, these fused proteins localized constitutively to actin filaments. While blue light illumination did not modify cytoplasmic localization of Arpin-CRY2 expressed alone, when it was co-expressed with LifeAct-CIBN or Utr-CIBN, blue light illumination induced its recruitment to actin filaments (Figure 5B; Movie S4). To simplify the analysis, cells were plated on thin line patterns, thus imposing a single protrusion (Doyle et al., 2009) and only two possible directions of migration. Continuous whole cell blue light illumination generated concentration gradients of Arpin-CRY2 from front to back of the cell, reflecting the binding of Arpin-CRY2 to actin and its advection with the retrograde flow (Figure 5C). Upon binding through Utrophin-CIBN the concentration profile rapidly changed to reach a new steady state within  $\sim 10$  min (Figure 5D). Different steady-state slopes were generated based on the association time of actin-binding proteins to actin, faint for LifeAct-CIBN, but much stronger for Utrophin-CIBN (Figure 5E), as expected from Figure 2F. This artificial system thus proved well suited to test our model: a negative regulator of actin polymerization, Arpin, was advected back from the cell leading edge by the actin retrograde flow, and the advection was more efficient for a stronger binding to actin.

We next assessed how this advection would affect cell migration. As expected, Arpin-CRY2 overexpression alone strongly impaired both cell speed and persistence (Dang et al., 2013), and co-expression of LifeAct-CIBN, which produces a weak backward advection of Arpin, was not able to rescue this effect after blue-light illumination. Conversely, co-expression of Utrophin-CIBN completely rescued cell speed and persistence time (Figures 5F and 5G), consistent with the different advection rates

of the Utrophin and Lifeact probes. To rule out that this effect was due to more efficient global depletion of Arpin from the cytoplasm upon binding to actin filaments via the Utrophin probe, we measured Arpin-CRY2 depletion from the cytoplasm after light-induced recruitment to the actin cytoskeleton via LifeAct-CIBN or Utrophin-CIBN. This showed that depletion was similar for both probes (Figure 5H). We could thus conclude that the rescue of cell speed and persistence observed with the Utrophin probe was specifically due to the advection of Arpin by the actin retrograde flow. This set of experiments provided two important validations of our model: first it showed that the effect of actin retrograde flow on cell persistence, via stabilization of a protrusion, which we observed in dendritic cells, could be reproduced in a completely different cell type, RPE1 cells. Second, it showed that direct engineering of an artificial system involving the minimal set of elements used in our model was enough to recapitulate the effects predicted by the model.

### Phase Diagram of Main Cell Migration Patterns

In addition to the prediction of the UCSP, our model provides through Equation 5 an explicit construction of a cell trajectory as that of an active Brownian particle (Romanczuk et al., 2012). While this concept has already proved to be useful to model phenomenologically cell trajectories (Selmeczi et al., 2008), so far no such bottom up approach was available. The analysis of Equation 5 yields a very rich phase diagram as a function of  $\beta$  and  $C_s$ , which predicts three main classes of trajectories, as detailed below (Figures 4B–4H and S5; see Supplemental Information for details).

#### Brownian Trajectories

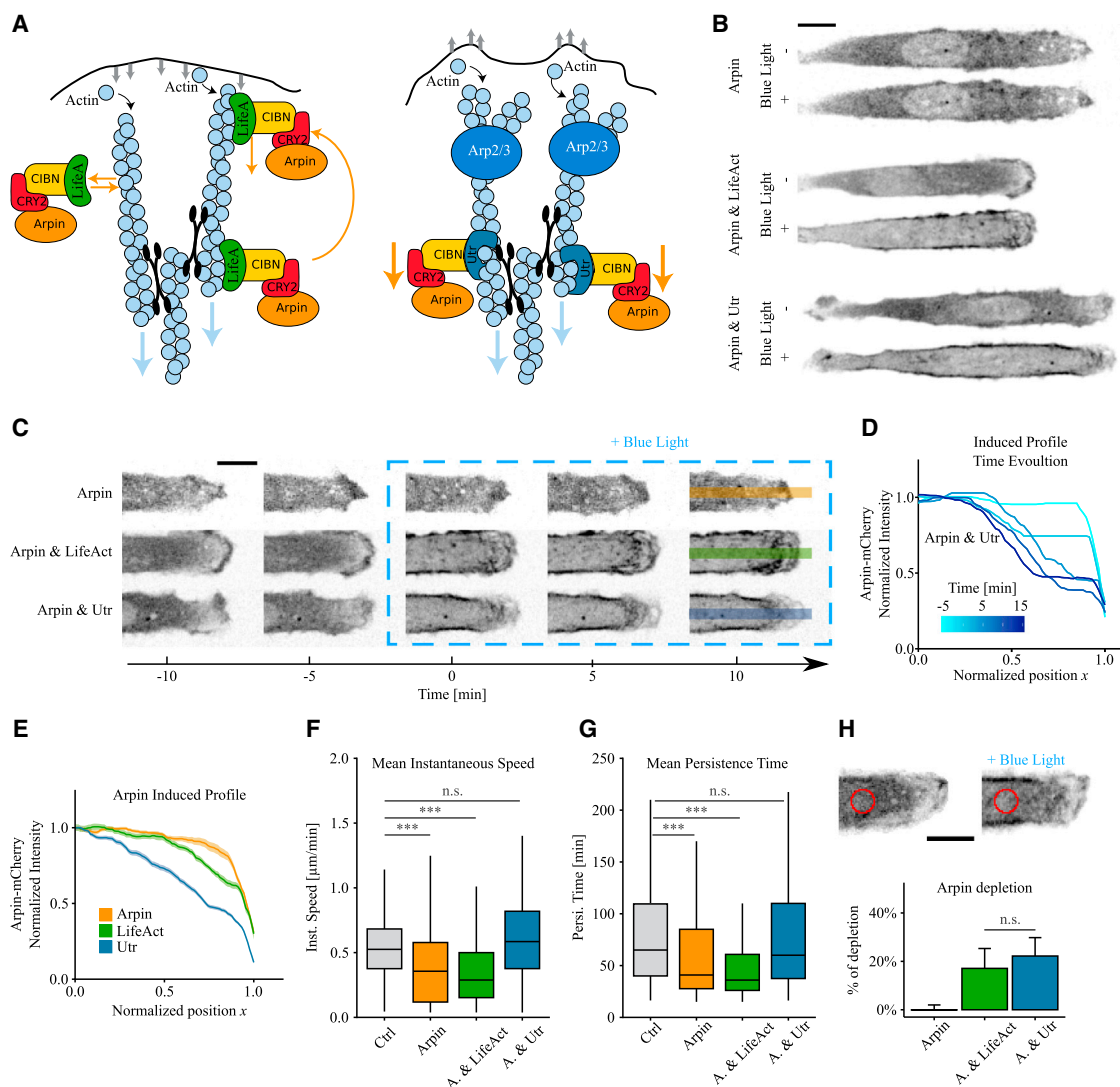
For  $\beta$  smaller than a critical value  $\beta_c(C_s)$ , the potential  $W(v)$  has a generic bowl shape centered at  $v = 0$  and the process can be well approximated by a classical Ornstein Uhlenbeck process (Gardiner, 2004) (Figures 4B, 4C, and 4F). This regime of slow maximal actin flow is characterized by an autocorrelation of  $v$  that decays exponentially with a short characteristic time  $\tau \sim 1/\gamma$ , so that there is no stable polarized state. At time scales larger than  $1/\gamma$  trajectories are then Brownian like.

#### Persistent Trajectories

For  $\beta > \beta_c(C_s)$  and  $C_s$  larger than a critical value  $C_s^c(\beta)$ ,  $W(v)$  has a “sombbrero shape,” which is the hallmark of systems with broken symmetry (Figures 4B, 4D, and 4G). In this regime of fast maximal actin flow and large maximal concentration of activated cues, the modulus of the velocity  $v$  fluctuates around a non-zero average. Trajectories then correspond to a persistent random walk with long lived polarization time, which is exponentially larger than  $1/\gamma$ .

#### Intermittent Trajectories

For  $\beta > \beta_c(C_s)$  and  $C_s < C_s^c(\beta)$ ,  $W(v)$  as a mixed shape, with both a local minimum around  $v = 0$  and a secondary minimum for a non-zero value of  $v$  (Figures 4B, 4E, and 4H). This regime of fast maximal actin flow and small maximal concentration of activated cues leads to intermittent trajectories (Bénichou et al., 2011), characterized by an alternation of Brownian and persistent phases. The stabilization of the Brownian phase (around  $v = 0$ ) is due to the multiplicative noise term in Equation 5 (see Mallick and Marcq, 2004): the small maximal concentration of activated cues induces large fluctuations that are enhanced for large actin



**Figure 5. Optically Triggered Advection of Arpin by the Actin Retrograde Flow**

(A) Schematic of the principle of the experiment: Arpin-CRY2 overexpression inhibits Arp2/3-dependent actin polymerization. Blue light illumination induces CIBN/CRY2 heterodimerization and the subsequent transient binding of Arpin-CRY2 to the co-expressed actin binding protein, LifeAct-CIBN (left) or Utr-CIBN (right). Binding to Utr-CIBN, but not binding to LifeAct-CIBN, induces depletion of Arpin from the leading edge.

(B–H) RPE1 cells plated on 9- $\mu\text{m}$  wide fibronectin-coated line micro-patterns, transiently transfected with three sets of constructs: Arpin-CRY2-mCherry alone, Arpin-CRY2-mCherry and LifeAct-CIBN(-GFP), or Arpin-CRY2-mCherry and Utr-CIBN(-GFP). (B) Confocal images of Arpin-CRY2-mCherry before (–) and after (+) blue light illumination in representative cells expressing the three sets of constructs. (C and D) Time-lapse images of the leading edge of the cells shown in (B).

(D) Time evolution of Arpin-CRY2-mCherry fluorescence intensity profile after light-induced binding to Utr-CIBN in the leading edge of a representative cell. Time is color-coded. In (D) and (E), the normalized position  $x$  is defined along the cell polarity axis, where the cell front is set at  $x = 1.0$ .

(E) Average steady state and SE of normalized Arpin-CRY2-mCherry fluorescence intensity profiles after blue light illumination at the leading edge of moving cells transfected with the three sets of constructs. (F and G) Instantaneous speed (F) and persistence time (G) of non-transfected cells (Ctrl) or cells transfected with the three sets of constructs. In boxplots: middle bars are medians, the rectangles span from the first to the third quartiles and bars extent from  $\pm 1.5 \times \text{IQR}$ . One-tailed  $t$  test.

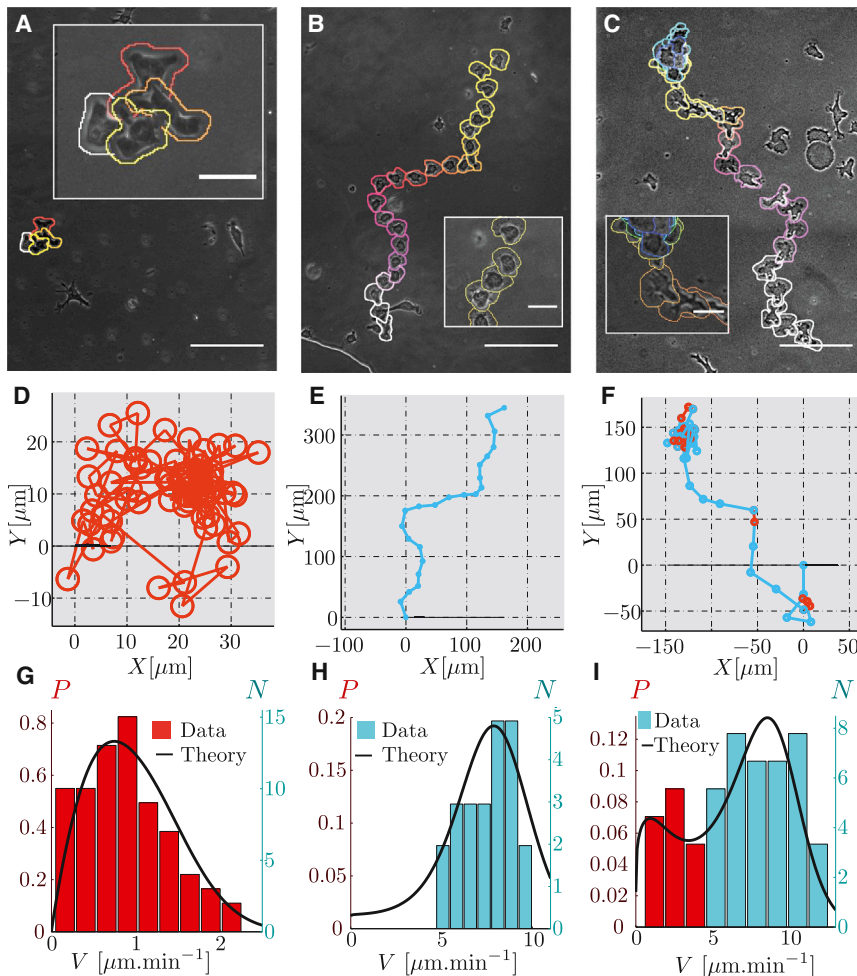
(H) Arpin-CRY2-mCherry depletion from the cytoplasm. Fluorescence was measured in the cytoplasm away from actin rich regions (red circle, top panel) for cells expressing the three sets of constructs (two-tailed  $t$  test, error bars are SD). \*\*\* $p < 0.001$ , n.s. not significant.

See also [Movie S4](#).

flows. The resulting dependence on  $v$  of the noise intensity leads to an effective restoring force toward the unpolarized state  $v = 0$ .

The analysis above shows in particular that the general coupling that we describe between actin flow and polarity cues is in principle sufficient to induce cell polarization (see the transi-

tion from diffusive to persistent or intermittent trajectories for increasing  $\beta$ ). We, however, do not claim that this is the only mechanism responsible for cell polarization. In fact, a preexisting polarization mechanism, for example of Turing type, can be included in the model (see [Supplemental Information](#)). In such



**Figure 6. Experimental Cell Trajectories Can Be Classified in the Three Classes Predicted by the Model**

Experimental cell trajectories can be classified in the three classes predicted by the model: diffusive (A, D, and G), persistent (B, E, and H), and intermittent (C, F, and I).

(A–C) Examples of BMDCs migration patterns of each type (temporal overlay of phase contrast images). Color code indicates the time course (total duration: (A) 276 min, (B) 72 min, and (C) 141 min. Scale bar, 100  $\mu\text{m}$  (insets: 25  $\mu\text{m}$ ).

(D–F) Corresponding trajectories extracted from automated tracking of the nucleus. Circles indicate the confidence interval (3  $\mu\text{m}$ ). Blue stands for cell speed  $v > 4 \mu\text{m}\cdot\text{min}^{-1}$  and red for  $v < 4 \mu\text{m}\cdot\text{min}^{-1}$ .

(G–I) Histograms of velocities extracted from the corresponding experimental tracks are in agreement with the distribution of velocities  $P(v)$  (solid black line) from the model with parameters determined in the [Supplemental Information](#) and  $\beta, C_s$  as indicated in [Figure 4B](#): diffusive phase (+), persistent (\*), and intermittent (x). See also [Movie S5](#).

a case, we found that the coupling to actin flows again results in an increased polarization lifetime and thus an increase in cell persistence. This shows the validity of the UCSP independently of the presence of a preexisting polarization mechanism.

The three classes of trajectories predicted by the model have been reported repeatedly in the literature ([Selmeczi et al., 2008](#); [Vedel et al., 2013](#)), which provides a further validation of the model. To test this prediction more quantitatively, we analyzed 2D trajectories of BMDCs ([Figures 6A–6F](#); [Movie S5](#)) obtained by automated tracking of cell nuclei. For each analyzed trajectory, the measured velocity distribution  $P(v)$  could be well fitted by the model by adjusting only  $\beta$  and  $C_s$ , while all other parameters values were kept as in [Figure 4](#) ([Figures 6G–6I](#)). We found that indeed all trajectories could be classified according to the above 3 classes predicted by the model (Brownian, persistent, intermittent) depending on the value of the parameters  $\beta$  and  $C_s$  only, in agreement with the predicted phase diagram ([Figure 4B](#)).

## DISCUSSION

Maintenance of cell polarity in time determines long-term cell migration patterns. Different molecular regulators of cell polarity

in motile cells have been identified in the past. However, mechanistic models with predictive capacity from molecular scales to global cell behavior (e.g., cell shape, speed, and persistence) remain sparse mainly due to a limitation of experimentally accessible parameters and the large number of involved components. In this work, we describe a positive feedback loop between actin flows and maintenance of cell polarity in motile cells.

This positive feedback impacts on the long-term migration behavior of cells and results in a higher persistence for faster cells—the UCSP law. By combining experimental results with theoretical modeling we suggest that actin flows are involved in reinforcing an asymmetric molecular distribution during cell polarization and thus increasing polarization lifetime. The two main predictions of the model were validated experimentally: (1) we first showed that it quantitatively predicts the UCSP law, and (2) we next demonstrated that it reproduces the main migratory behaviors reported so far. In addition, we could engineer a synthetic module based on the main ingredients of the model, to optically modulate the UCSP. We thus believe that our model provides an important step toward a comprehensive view on cell motility from microscopic to large-scale cell behavior.

The persistence time depends exponentially on the mean instantaneous velocity—that characterizes the UCSP—([Figure 1](#)) in the lower range of speeds for each cell type, and a saturation to a plateau is observed at larger speeds. Such saturation could have several causes. The model primarily predicts an exponential dependence of the polarization time  $\tau_p$  on the actin flow speed  $V$ . Such dependence  $\tau_p(V)$  implies a similar exponential dependence  $\tau(v)$  between persistence time and cell speed



under two conditions. First, it requires that polarization time  $\tau_p$  and persistence time  $\tau$  can be identified. While this is clear in 1D geometries, this does not always hold in dimensions 2 and 3, where two effects compete to destroy persistence: depolarization, characterized by the timescale  $\tau_p$ , and angular diffusion, characterized by the timescale  $\tau_\phi \sim \beta^2/K$ . One therefore expects that  $\tau \sim \tau_p$  only for  $\tau_p \leq \tau_\phi$ , while  $\tau$  saturates for larger values of  $\tau_p$ . Second, it requires that cell speed and actin flow speed are linearly coupled ( $v = \alpha V$ ). As we argued, such linear dependence generally holds in the lower range of speeds for each cell type. Non-linear effects (that could be due for example to a motor-clutch mechanism) (Mitchison and Kirschner, 1988) are, however, expected at larger speeds and could result in the observed saturation of the persistence time.

An important observation we made is that this law also applies at the subcellular scale, to individual cell protrusions. This has two consequences: first it explains why cells with various modes of migration follow the UCSP law, as it might apply to any locomotory subpart of the cell and the nature of the protrusions does not matter. Second, it suggests that such coupling between actin filaments flow rate (or even flows of other cytoskeletal elements) and lifetime of polarity might apply to other phenomena than cell migration, such as polarized secretion or growth, which also rely on actin polymerization.

In this work, we have validated the law on cells of mammalian origin migrating in 1D and 2D geometries, with or without confinement and in 3D collagen gels—the main *in vitro* migration assays. We also extended our finding to myeloid cells moving in live tissues in Medaka fish. The process we describe is so generic, that it is likely to apply to cells from other organisms. Indeed a very similar correlation between speed and persistence was reported from migrating amoeba (Miyoshi et al., 2003; Golé et al., 2011). This law *a priori* only applies to random migration and not to guided migration. It is, however, likely that the mechanism that we describe helps reinforce a weak external guidance and it might thus be also important for guided migration. As a first modeling step in that direction, we showed that our model and the UCSP still hold if an independent polarization module exists.

In this work, we did not aim at investigating in more details the molecules that might, for a given cell type, be responsible for coupling the flow to the polarity. We believe that, even in a given cell type, there might be several molecules that could contribute to the coupling. For example, one of the most obvious player would be Myosin II, as it fits the two requirements: it is transported by actin filaments and the steepest its gradient, the faster the flow. Such a mechanism inspired us in the choice of the feedback equation we used in the model (Equation 3). This hypothesis was validated in the case of BMDCs, which are typical amoeboid cells whose motility strongly relies on the activity of Myosin II. Nevertheless, inhibiting Myosin II was also shown to induce a stronger Rac-dependent protrusive activity and thus a stronger polymerization-based actin flow (Sanz-Moreno et al., 2008), which can also transport polarity cues and thus compensate for the loss of Myosin II, at least in protrusion-based migration in mesenchymal cells. Last, we stress that our analysis does not exclude polarity cues involved in the regulation of other key actors of cell polarity such as microtubules (Zhang et al., 2014), as long as they interact at least indirectly with actin.

One of the most promising aspect of the model we have proposed based on the UCSP, is that it can generate all the range of observed cell trajectories, with only two main parameters: the maximal actin flow velocity  $\beta$  and the maximal concentration of activated cues  $C_s$ . While the latter might be difficult to vary experimentally, we have shown that the former is in fact quite versatile (Figure 2A), which opens the way to the control of cell migration patterns. We could experimentally implement an optogenetic synthetic module that performs that task by inducing binding of a negative regulator of actin polymerization (Arpin) to actin filaments. Upon blue light illumination, Arpin was coupled to Utrophin and thus depleted from the migration front by retrograde advection with the actin flow, which in turn stabilized the protrusion. It is possible that some cells regulate binding of polarity factors to actin to tune their persistence. Making use of recent results that identified optimal search patterns for both persistent (Tejedor et al., 2012) and intermittent trajectories (Bénichou et al., 2011), we anticipate that the search efficiency for a target (e.g., DCs searching for antigens in a tissue) (Heuzé et al., 2013) could be optimized by tuning  $\beta$  and  $C_s$ . The model we propose thus provides a very generic ingredient of cell migration that could be used as a basis to model any process in which individual cell trajectories matter, such as search processes by immune cells (Harris et al., 2012), neuronal cells migration, or invasion by cancer cells.

## EXPERIMENTAL PROCEDURES

In brief (see the Extended Experimental Procedures for a detailed description of the methods), bone marrow-derived dendritic cells were generated from the bone marrow extracted from mice and cultured as previously described. Human retinal pigment epithelial (RPE1, Clontech) were grown in standard conditions. Medaka fish: (*Oryzias latipes*) stocks were maintained as previously described. *Itgb2*<sup>-/-</sup>, *Itgb2*<sup>-/-</sup> LifactGFP, and wild-type mice were kept on a C57BL/6 background and bred in a conventional animal facility at IST Austria according to local regulations.

For 1D migration assays, RPE1 cells were placed on line-shaped micro-patterns coated with fibronectin, and BMDCs into fibronectin-coated microchannels. For 2D migration, cells were either regularly plated on a fibronectin-coated surface, or plated between two surfaces separated by micro-fabricated spacers to introduce confinement. To obtain large cell tracking area, nine adjacent fields of views were recorded and images stacks were stitched by a custom written software. For 3D migration, collagen gels were prepared using 1.6 mg/ml of bovine collagen (PureCol, Advanced BioMatrix). BMDCs were embedded in polymerizing collagen and then immediately confined between two glass surfaces spaced 5  $\mu$ m apart.

Medaka myeloid cells live imaging was performed on fish at 9–11 days post-fertilization mounted as previously reported. Under-agarose-assays were performed with cells inserted between an agar gel and a Petri dish, as previously described. Blebbistatin (Sigma) was used in a final concentration of 10–20  $\mu$ M. ML-141 (Sigma) was used at a final concentration of 20  $\mu$ M. BMDCs and RPE1 cells were transfected using commercial transfection kit (from Amaxa and Roche or Invitrogen, respectively) according to the manufacturers recommendations. Nuclei were stained with Hoechst.

Time-lapse movies of cell nuclei in 1D, 2D, 3D, and *in vivo*, were analyzed by a custom written program as previously described. Medaka myeloid cells and BMDCs under-agarose were manually tracked with Fiji. In 1D, the persistent time is defined as the time a cell moves in the same direction, in 2D, 3D, and *in vivo*, as the time it takes for the cell to change its initial direction by 90°. The differential angle is the angle between two consecutive displacements of a cell. BMDCs under-agarose tracks were analyzed using in-house algorithms implemented in MATLAB (MathWorks). The mean-square displacement (MSD) was fitted according to the Furths formula in order to extract the

persistence time. Actin dynamics were analyzed by kymographs as previously described. Protein relocalization was quantified using Fiji. For the analysis of polarization lifetimes migrating cells were imaged every 10 s and the number of protrusions were recorded and saved as vectors.

RPE1 cells transfected with Arpin-CRY2-mCherry were imaged with an inverted spinning disk confocal microscope (Roper/Nikon) with 60× magnification. mCherry channel was acquired every minute, then after 10 min, Arpin-CRY2-mCherry recruitment to F-actin (using actin-bound LifeAct-CIBN-GFP or Utrophin-CIBN-GFP) was induced by illuminating also in the GFP channel. For statistical analysis of trajectories, cells were imaged at lower magnification (5×) with illumination in the GFP channel at each time point (time-lapse 5 min).

Medaka fish (kindly provided by J. Wittbrodt) were kept and treated in accordance with the German (Tierschutzgesetz) or Italian (decree 116/92) national guidelines and experimental procedures were approved by Institutional Animal Care and Use Committee. Mice were bred in conventional animal facilities at IST Austria or at the Institut Curie according to local regulations.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, Supplemental Discussion, five movies, five figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.056>.

## AUTHOR CONTRIBUTIONS

P.M. was in charge of the project and performed the cell trajectory analysis. J.F.R. was the main contributor for the modeling and performed the simulations. S.W. performed and analyzed the experiments on dendritic cells presented in Figures 2 and 3. O.B. and N.G. contributed to the modeling. V.R. contributed to the experiments and analysis on dendritic cells presented in Figures 2 and 3. P.M., N.C., M.C., and S.D.B. designed the optogenetic tools and performed experiments of Figure 5. F.L., M.L.B., M.R., and H.R.T. performed cell migration experiments presented in Figures 1 and 6. C.L.C. performed the experiments on myeloid cells in live Medaka fish presented in Figure 1. C.P.H. supervised the work of V.R. A.-M.L.-D. supervised the experiments on dendritic cells presented in Figure 1. M.S. supervised the work of S.W. and contributed to the writing of the article. R.V. supervised the modeling, the work of J.F.R., and wrote the article. M.P. coordinated the study, supervised the work of P.M., F.L., M.L.B., M.R., and H.R.T., and wrote the article.

## ACKNOWLEDGMENTS

We acknowledge J. Wittbrodt for providing the medaka fish transgenic line and D. Riveline and J.F. Joanny for discussions and comments on the manuscript. The authors greatly acknowledge Lucie Sengmanivong from the Nikon Imaging Centre at Institut Curie-CNRS and Vincent Fraiser from the PICT-IBISA Lhomond Imaging facility of Institut Curie for their help on microscopy. P.M. was supported by a “Fondation pour la Recherche Médicale” fellowship. V.R. was supported by the Austrian Science Fund (FWF) T560-B17. C.L.C. was supported by grant AIRC-IG and PRIN 2010-2011 to Ruggero Pardi. F.L. was supported by EMBO ALTF 1163-2010. O.B., R.V., and M.P. were supported by ANR (ANR-09-PIRI-0027) and InNaBioSant. M.P. was supported by ERC (311205-PROMICO). M.S. was supported by ERC (281556\_Leukocyte Forces) and a HFSP Program Grant.

Received: February 4, 2014

Revised: November 25, 2014

Accepted: January 22, 2015

Published: March 19, 2015

## REFERENCES

Bénichou, O., Loverdo, C., Moreau, M., and Voituriez, R. (2011). Intermittent search strategies. *Rev. Mod. Phys.* 83, 81–129.

Bois, J.S., Jülicher, F., and Grill, S.W. (2011). Pattern formation in active fluids. *Phys. Rev. Lett.* 106, 028103.

Burkel, B.M., von Dassow, G., and Bement, W.M. (2007). Versatile fluorescent probes for actin filaments based on the actin-binding domain of utrophin. *Cell Motil. Cytoskeleton* 64, 822–832.

Callan-Jones, A.C., and Voituriez, R. (2013). Active gel model of amoeboid cell motility. *New J. Phys.* 15, 025022.

Condamine, S., Bénichou, O., Tejedor, V., Voituriez, R., and Klafter, J. (2007). First-passage times in complex scale-invariant media. *Nature* 450, 77–80.

Czirók, A., Schlett, K., Madarász, E., and Vicsek, T. (1998). Exponential distribution of locomotion activity in cell cultures. *Phys. Rev. Lett.* 81, 3038–3041.

Dang, I., Gorelik, R., Sousa-Blin, C., Derivery, E., Guérin, C., Linkner, J., Nemethova, M., Dumortier, J.G., Giger, F.A., Chipysheva, T.A., et al. (2013). Inhibitory signalling to the Arp2/3 complex steers cell migration. *Nature* 503, 281–284.

Doyle, A.D., Wang, F.W., Matsumoto, K., and Yamada, K.M. (2009). One-dimensional topography underlies three-dimensional fibrillar cell migration. *J. Cell Biol.* 184, 481–490.

Gardel, M.L., Schneider, I.C., Aratyn-Schaus, Y., and Waterman, C.M. (2010). Mechanical integration of actin and adhesion dynamics in cell migration. *Annu. Rev. Cell Dev. Biol.* 26, 315–333.

Gardiner, C. (2004). *Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences* (Berlin: Springer).

Goehring, N.W., Trong, P.K., Bois, J.S., Chowdhury, D., Nicola, E.M., Hyman, A.A., and Grill, S.W. (2011). Polarization of PAR proteins by advective triggering of a pattern-forming system. *Science* 334, 1137–1141.

Golé, L., Rivière, C., Hayakawa, Y., and Rieu, J.-P. (2011). A quorum-sensing factor in vegetative Dictyostelium discoideum cells revealed by quantitative migration analysis. *PLoS ONE* 6, e26901.

Grabher, C., Cliffe, A., Miura, K., Hayflick, J., Pepperkok, R., Rørth, P., and Wittbrodt, J. (2007). Birth and life of tissue macrophages and their migration in embryogenesis and inflammation in medaka. *J. Leukoc. Biol.* 81, 263–271.

Harris, T.H., Banigan, E.J., Christian, D.A., Konradt, C., Tait Wojno, E.D., Norose, K., Wilson, E.H., John, B., Weninger, W., Luster, A.D., et al. (2012). Generalized Lévy walks and the role of chemokines in migration of effector CD8+ T cells. *Nature* 486, 545–548.

Hawkins, R.J., Poincloux, R., Bénichou, O., Piel, M., Chavrier, P., and Voituriez, R. (2011). Spontaneous contractility-mediated cortical flow generates cell migration in three-dimensional environments. *Biophys. J.* 101, 1041–1045.

Heasman, S.J., and Ridley, A.J. (2008). Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nat. Rev. Mol. Cell Biol.* 9, 690–701.

Heuzé, M.L., Vargas, P., Chabaud, M., Le Berre, M., Liu, Y.-J., Collin, O., Solanes, P., Voituriez, R., Piel, M., and Lennon-Duménil, A.-M. (2013). Migration of dendritic cells: physical principles, molecular mechanisms, and functional implications. *Immunol. Rev.* 256, 240–254.

Hong, L., Kenney, S.R., Phillips, G.K., Simpson, D., Schroeder, C.E., Nöth, J., Romero, E., Swanson, S., Waller, A., Strouse, J.J., et al. (2013). Characterization of a Cdc42 protein inhibitor and its use as a molecular probe. *J. Biol. Chem.* 288, 8531–8543.

Jülicher, F., Kruse, K., Prost, J., and Joanny, J.F. (2007). Active behavior of the cytoskeleton. *Phys. Rep.* 449, 3–28.

Jurado, C., Haserick, J.R., and Lee, J. (2005). Slipping or gripping? Fluorescent speckle microscopy in fish keratocytes reveals two different mechanisms for generating a retrograde flow of actin. *Mol. Biol. Cell* 16, 507–518.

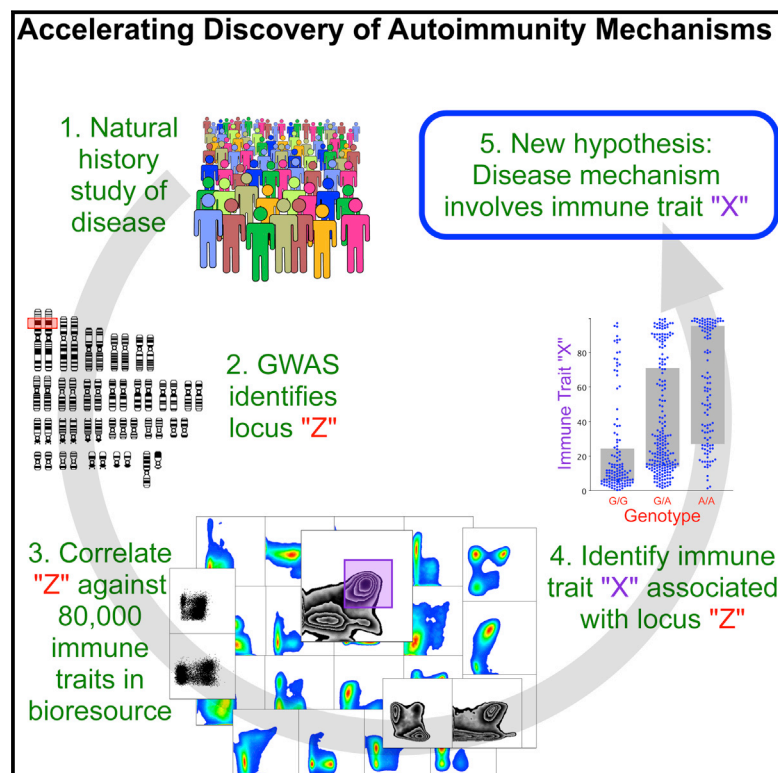
Kennedy, M.J., Hughes, R.M., Peteya, L.A., Schwartz, J.W., Ehlers, M.D., and Tucker, C.L. (2010). Rapid blue-light-mediated induction of protein interactions in living cells. *Nat. Methods* 7, 973–975.

Kondo, T., Hamao, K., Kamijo, K., Kimura, H., Morita, M., Takahashi, M., and Hosoya, H. (2011). Enhancement of myosin II/actin turnover at the contractile ring induces slower furrowing in dividing HeLa cells. *Biochem. J.* 435, 569–576.

- Lämmermann, T., Renkawitz, J., Wu, X., Hirsch, K., Brakebusch, C., and Sixt, M. (2009). Cdc42-dependent leading edge coordination is essential for interstitial dendritic cell migration. *Blood* 113, 5703–5710.
- Maiuri, P., Terriac, E., Paul-Gilloteaux, P., Vignaud, T., McNally, K., Onuffer, J., Thorn, K., Nguyen, P.A., Georgoulia, N., Soong, D., et al. (2012). The first world cell race. *Curr. Biol.* 22, R673–R675.
- Mallick, K., and Marcq, P. (2004). Noise-induced reentrant transition of the stochastic Duffing oscillator. *Eur. Phys. J B* 38, 99–102.
- Mitchison, T., and Kirschner, M. (1988). Cytoskeletal dynamics and nerve growth. *Neuron* 1, 761–772.
- Miyoshi, H., Masaki, N., and Tsuchiya, Y. (2003). Characteristics of trajectory in the migration of *Amoeba proteus*. *Protoplasma* 222, 175–181.
- Mogilner, A., and Oster, G. (1996). Cell motility driven by actin polymerization. *Biophys. J.* 71, 3030–3045.
- Munro, E., Nance, J., and Priess, J.R. (2004). Cortical flows powered by asymmetrical contraction transport PAR proteins to establish and maintain anterior-posterior polarity in the early *C. elegans* embryo. *Dev. Cell* 7, 413–424.
- Ofer, N., Mogilner, A., and Keren, K. (2011). Actin disassembly clock determines shape and speed of lamellipodial fragments. *Proc. Natl. Acad. Sci. USA* 108, 20394–20399.
- Poincloux, R., Collin, O., Lizárraga, F., Romao, M., Debray, M., Piel, M., and Chavrier, P. (2011). Contractility of the cell rear drives invasion of breast tumor cells in 3D Matrigel. *Proc. Natl. Acad. Sci. USA* 108, 1943–1948.
- Renkawitz, J., Schumann, K., Weber, M., Lämmermann, T., Pflicke, H., Piel, M., Polleux, J., Spatz, J.P., and Sixt, M. (2009). Adaptive force transmission in amoeboid cell migration. *Nat. Cell Biol.* 11, 1438–1443.
- Riedl, J., Crevenna, A.H., Kessenbrock, K., Yu, J.H., Neukirchen, D., Bista, M., Bradke, F., Jenne, D., Holak, T.A., Werb, Z., et al. (2008). Lifeact: a versatile marker to visualize F-actin. *Nat. Methods* 5, 605–607.
- Romanczuk, P., Bar, M., Ebeling, W., Lindner, B., and Schimansky-Geier, L. (2012). Active Brownian particles from individual to collective stochastic dynamics. *Eur. Phys. J. Special Topics* 202, 1–162.
- Sanz-Moreno, V., Gadea, G., Ahn, J., Paterson, H., Marra, P., Pinner, S., Sahai, E., and Marshall, C.J. (2008). Rac activation and inactivation control plasticity of tumor cell movement. *Cell* 135, 510–523.
- Selmeczi, D., Li, L., Pedersen, L.I.I., Norrelykke, S.F., Hagedorn, P.H., Mosler, S., Larsen, N.B., Cox, E.C., and Flyvbjerg, H. (2008). Cell motility as random motion: a review. *Eur. Phys. J. Special Topics* 157, 1–15.
- Svitkina, T.M., Verkhovsky, A.B., McQuade, K.M., and Borisy, G.G. (1997). Analysis of the actin-myosin II system in fish epidermal keratocytes: mechanism of cell body translocation. *J. Cell Biol.* 139, 397–415.
- Tejedor, V., Voituriez, R., and Bénichou, O. (2012). Optimizing persistent random searches. *Phys. Rev. Lett.* 108, 088103.
- Theriot, J.A., and Mitchison, T.J. (1991). Actin microfilament dynamics in locomoting cells. *Nature* 352, 126–131.
- Toner, J., Tu, Y., and Ramaswamy, S. (2005). Hydrodynamics and phases of flocks. *Ann. Phys.* 318, 170–244.
- Vedel, S., Tay, S., Johnston, D.M., Bruus, H., and Quake, S.R. (2013). Migration of cells in a social context. *Proc. Natl. Acad. Sci. USA* 110, 129–134.
- Vicente-Manzanares, M., Ma, X., Adelstein, R.S., and Horwitz, A.R. (2009). Non-muscle myosin II takes centre stage in cell adhesion and migration. *Nat. Rev. Mol. Cell Biol.* 10, 778–790.
- Wilson, C.A., Tsuchida, M.A., Allen, G.M., Barnhart, E.L., Applegate, K.T., Yam, P.T., Ji, L., Keren, K., Danuser, G., and Theriot, J.A. (2010). Myosin II contributes to cell-scale actin network treadmill through network disassembly. *Nature* 465, 373–377.
- Wu, P.-H., Giri, A., Sun, S.X., and Wirtz, D. (2014). Three-dimensional cell migration does not follow a random walk. *Proc. Natl. Acad. Sci. USA* 111, 3949–3954.
- Zhang, J., Guo, W.-H., and Wang, Y.-L. (2014). Microtubules stabilize cell polarity by localizing rear signals. *Proc. Natl. Acad. Sci. USA* 111, 16383–16388.

# The Genetic Architecture of the Human Immune System: A Bioresource for Autoimmunity and Disease Pathogenesis

## Graphical Abstract



## Authors

Mario Roederer, Lydia Quaye, ..., Tim D. Spector, Frank O. Nestle

## Correspondence

roederer@nih.gov (M.R.), tim.spector@kcl.ac.uk (T.D.S.)

## In Brief

The study of a large and homogenous population of human twins identifies numerous genetic loci controlling the phenotype or number of functionally important immune subsets in the blood, providing a database to test associations of any genetic locus with more than 78,000 different immune traits.

## Highlights

- Resource of heritabilities and genetic associations of 80,000 immune traits in 669 twins
- Genetic associations with immune cell frequencies and surface protein expression levels
- Of the top 150 traits, 11 genetic loci explained up to 36% of variation of 19 traits
- Loci include autoimmune susceptibility genes, providing etiological hypotheses





# The Genetic Architecture of the Human Immune System: A Bioresource for Autoimmunity and Disease Pathogenesis

Mario Roederer,<sup>1,7,\*</sup> Lydia Quaye,<sup>2,7</sup> Massimo Mangino,<sup>2,4,7</sup> Margaret H. Beddall,<sup>1</sup> Yolanda Mahnke,<sup>1,5</sup> Pratip Chattopadhyay,<sup>1</sup> Isabella Tosi,<sup>3,4</sup> Luca Napolitano,<sup>3</sup> Manuela Terranova Barberio,<sup>3</sup> Cristina Menni,<sup>2</sup> Federica Villanova,<sup>3,4</sup> Paola Di Meglio,<sup>3,6</sup> Tim D. Spector,<sup>2,8,\*</sup> and Frank O. Nestle<sup>3,4,8</sup>

<sup>1</sup>ImmunoTechnology Section, Vaccine Research Center, NIAID, NIH, Bethesda, MD 20892, USA

<sup>2</sup>Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH, UK

<sup>3</sup>Cutaneous Medicine Unit, St. John's Institute of Dermatology, King's College London, London SE1 9RT, UK

<sup>4</sup>NIHR Biomedical Research Centre at Guy's and St. Thomas' NHS Foundation Trust, London SE1 9RT, UK

<sup>5</sup>Present address: Translational and Correlative Sciences Laboratory, Translational Research Program, Perelman School of Medicine, The University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>6</sup>Present address: Division of Molecular Immunology, MRC National Institute for Medical Research, Mill Hill, London NW1 7AA, UK

<sup>7</sup>Co-first author

<sup>8</sup>Co-senior author

\*Correspondence: [roederer@nih.gov](mailto:roederer@nih.gov) (M.R.), [tim.spector@kcl.ac.uk](mailto:tim.spector@kcl.ac.uk) (T.D.S.)

<http://dx.doi.org/10.1016/j.cell.2015.02.046>

## SUMMARY

Despite recent discoveries of genetic variants associated with autoimmunity and infection, genetic control of the human immune system during homeostasis is poorly understood. We undertook a comprehensive immunophenotyping approach, analyzing 78,000 immune traits in 669 female twins. From the top 151 heritable traits (up to 96% heritable), we used replicated GWAS to obtain 297 SNP associations at 11 genetic loci, explaining up to 36% of the variation of 19 traits. We found multiple associations with canonical traits of all major immune cell subsets and uncovered insights into genetic control for regulatory T cells. This data set also revealed traits associated with loci known to confer autoimmune susceptibility, providing mechanistic hypotheses linking immune traits with the etiology of disease. Our data establish a bioresource that links genetic control elements associated with normal immune traits to common autoimmune and infectious diseases, providing a shortcut to identifying potential mechanisms of immune-related diseases.

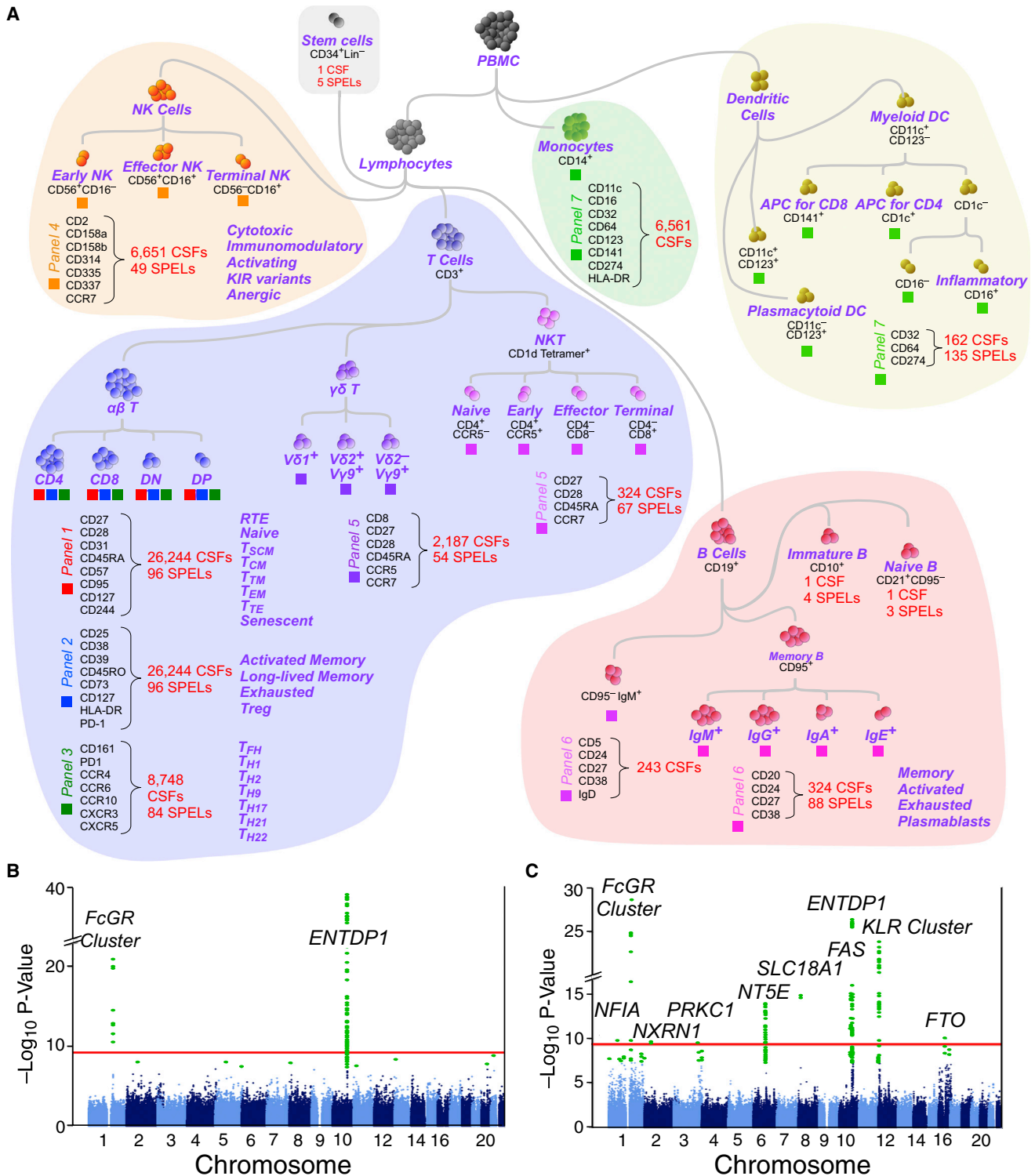
## INTRODUCTION

The immune system has evolved over millions of years into a remarkable defense mechanism with rapid and specific protection of the host from major environmental threats and pathogens. Such pathogen encounters have contributed to a selection of immune genes at the population level that determine not only host-specific pathogen responses but also susceptibility to autoimmune disease and immunopathogenesis. Understanding how such genes interplay with the environment to determine im-

mune protection and pathology is critical for unravelling the mechanisms of common autoimmune and infectious diseases and future development of vaccines and immunomodulatory therapies.

Studies of rare disease established major genes, and their associated pathways, that regulate pathogen-specific immune responses (Casanova and Abel, 2004) and genome-wide association studies (GWAS) of autoimmune disease have also been productive for finding common variants (Cotsapas and Hafler, 2013; Parkes et al., 2013; Raj et al., 2014). Despite this progress, there are still major limitations in our understanding of the genetics of complex autoimmune or infectious diseases. A key missing piece is the elucidation of the genes controlling critical components of a normal human immune system under homeostatic conditions. These include the relative frequencies of circulating immune cell subsets and the regulation of cell-surface expression of key proteins that we expect have strong regulatory mechanisms.

Previous studies in humans and rodents have shown that variation in the levels of circulating blood T cells is in part heritable (Amadori et al., 1995; Kraal et al., 1983). Identifying the underlying genetic elements would help us understand the mechanisms of homeostasis—and its dysregulation. Twin studies are ideal to quantify the heritability of immune traits in healthy humans by allowing adjustment for the influence of genes, early environment, age, and cohort, plus a number of known and unknown confounders (van Dongen et al., 2012). Early studies from our group demonstrated genetic control of CD8 and CD4 T cell levels in twins (Ahmadi et al., 2001), and others have shown similar heritable effects in non-twins and rodents and with broad white cell phenotypes (Amadori et al., 1995; Clementi et al., 1999; Damoiseaux et al., 1999; Evans et al., 1999; Ferreira et al., 2010; Hall et al., 2000; Kraal et al., 1983; Nalls et al., 2011; Okada et al., 2011). A recent study, with a family design, was the first to perform GWAS on a larger range of immune subtypes. The authors analyzed 272 correlated immune traits derived from 95



cell types and described 23 independent genetic variants within 13 independent loci (Orrù et al., 2013).

Here, we report a comprehensive and high-resolution deep immunophenotyping flow cytometry analysis in 669 female twins using 7 distinct 14-color immunophenotyping panels that captured nearly 80,000 cell types (comprising ~1,800 independent phenotypes) to analyze both immune cell subset frequency (CSF) and immune cell-surface protein expression levels (SPELs). This gave us a roughly 30-fold richer view of the healthy immune system than was previously achievable. Taking advantage of the twin model, we used a pre-specified analysis plan that prioritized 151 independent immune traits for genome-wide association analysis and replication.

We find 241 genome-wide significant SNPs within 11 genetic loci, 9 of which are previously unreported. Importantly, they explain up to 36% of the variation of 19 immune traits (18 previously unexplored). We identify pleiotropic “master” genetic loci controlling multiple immune traits and key immune traits under tight genetic control by multiple genetic loci. In addition, we show the importance of quantifying cell-surface antigen expression rather than just cell-type frequency.

Critically, we show overlap between these genetic associations of normal immune homeostasis with previously established autoimmune and infectious disease associations. This rich database provides a vital, publicly accessible bioresource as a bridge between genetic and immune discoveries that will expedite the identification of disease mechanisms in autoimmunity and infection.

## RESULTS

### Subjects

The discovery stage comprised 497 female participants from the UK Adult Twin Register (TwinsUK). There were 75 complete monozygotic (MZ) twin pairs, 170 dizygotic (DZ) pairs, and 7 singletons (arising from quality control [QC] failures in one co-twin). The mean age was 61.4 years (range: 40–77). The replication stage comprised a further 172 participants, mean age 58.2 years (range: 32–83), with 46 MZ, 118 DZ, and 8 singletons. We stained cryopreserved peripheral blood mononuclear cells (PBMC) from each, using a set of 7 14-color immunophenotyping panels that delineate a large range of immune subsets (Figures 1A, S1, S2, and S3 and Table S1). Immune traits analyzed included the CSF (i.e., the proportionate representation of a given phenotype) and the SPEL (i.e., a quantitative measure of gene expression on a per-cell basis). The variability of all traits was assessed using longitudinal sampling on a small cohort of individuals as described in the Experimental Procedures; of the 50,000 traits meeting the first filter criterion (Figure S4), the mean covariance across samples drawn 6 months apart is 0.86. All trait values and summary analyses, including variability, are available for download. Data and statistical analysis of the discovery stage was completed per a pre-defined statistical

analysis plan before samples from the replication stage were thawed.

GWAS analysis of all 78,000 immune traits is computationally prohibitive and would require a multiple comparisons correction that dramatically reduces sensitivity. The ability to infer heritability (proportion of variance explained solely by genetic factors) by the use of twins dramatically enhanced our ability to focus on those that are most likely to be informative. Co-variation of all traits was computed; about 1,800 were independent at  $r < 0.7$  (Figure S4).

We found no significant association of the analyzed traits with self-reported tobacco use or alcohol consumption and so did not include those behaviors as covariates. We identified many traits associated with age and included age as a covariate in all analyses. Notably, an advantage of using a twin-based cohort is to render age and other cohort effects minimally impactful. The age range of our cohort was optimal for our goal of identifying immune traits associated with genetic elements that show a risk for autoimmune diseases. Because incidence for such diseases often increases with age, the greatest power for such correlations will be obtained using samples measurements most proximal to the common onset of disease.

### Heritability

Falconer's traditional formula (twice the difference in intraclass correlations) was used to roughly estimate the heritabilities of all 78,000 immune traits; after ranking, traits were selected for further pre-specified analyses (Figure S4). Variance components analysis (additive genetics, common environment, and unique environment, or ACE model) was used to more precisely estimate heritabilities of chosen traits. The heritabilities ranged widely from 0%—suggesting purely environmental or stochastic influences—to 96% (e.g., CD32 expression on dendritic cells), indicating a strong genetic effect. Figure S5 shows the range of heritabilities for selected traits, and the components of the model are tabulated in Table S2 with full trait descriptions.

### GWAS of Immune Traits

Single-variant associations were performed on 151 immune traits selected for high heritability or biological interest, comprising cell frequency (129 CSFs) and cell-surface protein expression (22 SPELs). Many significant associations were found despite the stringent Bonferroni multiple testing threshold of  $p < 3.3 \times 10^{-10}$ . We also performed a conditional analysis, including the top SNP of each locus as a covariate, to identify potential independent secondary signals. This analysis did not reveal any significant evidence for additional independent signals.

Six SPELs were significant (Table 1), with the strongest between MFI:516 (CD39 SPEL on CD4 T cells) and the *ENTPD1* (CD39 gene) SNP rs7096317 ( $p = 9.4 \times 10^{-40}$ ). Many other variants of *ENTPD1* were also associated with this trait (Table S3). Expression of five others (MFI:189, MFI:212, MFI:231, MFI:504, and MFI:552, which include CD27 expression on B and T cell

(B and C) Summary Manhattan plots: green dots, genome-wide significant associations ( $p < 5 \times 10^{-8}$ ). The red line indicates the significance threshold of  $p < 3.3 \times 10^{-10}$ , which corresponds to the standard genome-wide threshold after further adjustment for 151 independent tests. The variants shown are MAF  $\geq 0.1$ ; call rate  $\geq 0.9$ ; HWE  $p$  value  $\geq 1 \times 10^{-6}$ . Shown are separate plots for SPEL associations (B) and CSF associations (C).

**Table 1. Discovery and Replication Results for the Top Significant SNPs at Each Locus for Each Immune Trait**

Locus:Genes	Trait ID	Trait Phenotype	Marker	Chr	EA/NEA	EAF	Beta (SE)	p Value	Beta (SE)	p Value	Beta (SE)	p Value
1: <i>FCGR2A</i> , <i>FCGR2B</i> , <i>FCRLA</i>	MFI:189	CD27 on IgA <sup>+</sup> B	rs1801274	1	A/G	0.49	0.128 (0.02)	6.48E–11	0.07 (0.03)	3.70E–02	0.11 (0.02)	2.8E–11
1: <i>FCGR2A</i> , <i>FCGR2B</i> , <i>FCRLA</i>	MFI:212	CD27 on IgG <sup>+</sup> B	rs1801274	1	A/G	0.49	0.136 (0.02)	5.38E–12	0.12 (0.03)	1.11E–04	0.13 (0.02)	2.9E–15
1: <i>FCGR2A</i> , <i>FCGR2B</i> , <i>FCRLA</i>	MFI:231	CD161 on CD4 T	rs1801274	1	A/G	0.49	0.131 (0.02)	2.64E–11	0.12 (0.03)	2.17E–04	0.13 (0.02)	2.7E–14
1: <i>FCGR2A</i> , <i>FCGR2B</i> , <i>FCRLA</i>	MFI:504	CD27 on CD4 T	rs1801274	1	A/G	0.49	0.145 (0.02)	5.42E–14	0.14 (0.03)	4.20E–05	0.14 (0.02)	1.1E–17
1: <i>FCGR2A</i> , <i>FCGR2B</i> , <i>FCRLA</i>	MFI:552	CD27 on CD8 T	rs1801274	1	A/G	0.49	0.186 (0.02)	1.26E–21	0.12 (0.03)	1.72E–04	0.17 (0.02)	4.2E–24
1: <i>FCGR2A</i> , <i>FCGR2B</i> , <i>FCRLA</i>	P7:110	iMDC: %CD32 <sup>+</sup>	rs10494359	1	C/G	0.12	0.343 (0.03)	2.52E–29	0.43 (0.05)	1.05E–15	0.36 (0.03)	5.9E–43
1: <i>FCGR2A</i> , <i>FCGR2B</i> , <i>FCRLA</i>	P7:224	CD1c <sup>+</sup> mDC: %CD32	rs4657090	1	A/G	0.27	–0.174 (0.02)	1.30E–14	–0.19 (0.04)	3.86E–06	–0.18 (0.02)	2.7E–19
2: <i>NFIA</i>	P4:3551	NK: %CD314 <sup>–</sup> CD158a <sup>+</sup>	rs12072379	1	G/C	0.16	–0.131 (0.02)	1.73E–10	–0.10 (0.05)	4.87E–02	–0.13 (0.02)	2.7E–11
3: <i>NRXN1</i>	P4:3551	NK: %CD314 <sup>–</sup> CD158a <sup>+</sup>	rs17040907	2	T/C	0.07	–0.208 (0.03)	2.68E–10	–0.16 (0.08)	4.22E–02	–0.02 (0.03)	3.9E–11
4: <i>PRKCI</i>	P4:3551	NK: %CD314 <sup>–</sup> CD158a <sup>+</sup>	rs2650220	3	G/A	0.16	–0.15 (0.02)	3.18E–10	–0.10 (0.05)	4.55E–02	–0.14 (0.02)	6.0E–11
5: <i>NT5E</i>	P2:4195	CD4 T: %CD39–CD73 <sup>+</sup>	rs9444346	6	G/A	0.19	–0.2 (0.03)	1.18E–14	–0.12 (0.04)	4.98E–03	–0.18 (0.02)	8.8E–16
<i>RP11–30P6</i>	P2:4204	CD4 T: %CD73 <sup>+</sup>	rs9444346	6	G/A	0.19	–0.195 (0.03)	5.85E–14	–0.12 (0.04)	2.82E–03	–0.18 (0.02)	1.8E–15
6: <i>SLC18A1</i>	P4:3551	NK: %CD314 <sup>–</sup> CD158a <sup>+</sup>	rs1390942	8	T/C	0.15	–0.163 (0.02)	1.39E–15	–0.20 (0.05)	1.70E–04	–0.17 (0.02)	1.4E–18
7: <i>SLC25A16</i>	P4:3551	NK: %CD314 <sup>–</sup> CD158a <sup>+</sup>	rs3017072	10	T/C	0.15	–0.153 (0.02)	2.75E–13	–0.15 (0.05)	2.08E–03	–0.15 (0.02)	2.2E–15
8: <i>FAS</i> , <i>ACTA2</i>	P1:6601	CD8 T: %TSCM	rs7097572	10	C/T	0.48	–0.168 (0.02)	8.51E–16	–0.18 (0.03)	2.72E–06	–0.17 (0.02)	1.3E–20
9: <i>ALDH18A1</i> , <i>ENTPD1</i> , <i>ENTPD1–AS1</i> , <i>RP11–7D5</i> , <i>SORBS1</i> , <i>TCTN3</i>	MFI:516	CD39 on CD4 T	rs7096317	10	G/A	0.42	–0.255 (0.02)	9.40E–40	–0.30 (0.04)	9.92E–17	–0.27 (0.02)	1.6E–54
9: <i>ALDH18A1</i> , <i>ENTPD1</i> , <i>ENTPD1–AS1</i> , <i>RP11–7D5</i> , <i>SORBS1</i> , <i>TCTN3</i>	P2:10491	CD8 T: %CD39 <sup>+</sup>	rs4074424	10	G/C	0.42	–0.219 (0.02)	4.11E–27	–0.19 (0.04)	2.40E–07	–0.21 (0.02)	8.2E–33
9: <i>ALDH18A1</i> , <i>ENTPD1</i> , <i>ENTPD1–AS1</i> , <i>RP11–7D5</i> , <i>SORBS1</i> , <i>TCTN3</i>	P2:3460	CD4 T: %CD39 <sup>+</sup> CD38 <sup>+</sup> PD1 <sup>–</sup>	rs4582902	10	C/T	0.47	–0.164 (0.02)	4.55E–16	–0.19 (0.03)	2.58E–08	–0.17 (0.02)	9.2E–23
9: <i>ALDH18A1</i> , <i>ENTPD1</i> , <i>ENTPD1–AS1</i> , <i>RP11–7D5</i> , <i>SORBS1</i> , <i>TCTN3</i>	P2:4159	CD4 T: %CD39 <sup>+</sup> CD73 <sup>–</sup>	rs6584027	10	G/A	0.47	–0.212 (0.02)	1.54E–25	–0.20 (0.04)	1.09E–08	–0.21 (0.02)	1.1E–32
9: <i>ALDH18A1</i> , <i>ENTPD1</i> , <i>ENTPD1–AS1</i> , <i>RP11–7D5</i> , <i>SORBS1</i> , <i>TCTN3</i>	P2:4186	CD4 T: %CD39 <sup>+</sup>	rs6584027	10	G/A	0.47	–0.215 (0.02)	2.20E–26	–0.21 (0.04)	5.27E–09	–0.21 (0.02)	7.8E–34
9: <i>ALDH18A1</i> , <i>ENTPD1</i> , <i>ENTPD1–AS1</i> , <i>RP11–7D5</i> , <i>SORBS1</i> , <i>TCTN3</i>	P2:4213	CD4 T: %CD39 <sup>+</sup> CD73 <sup>+</sup>	rs10882676	10	A/C	0.47	–0.195 (0.02)	6.76E–22	–0.19 (0.04)	2.19E–07	–0.19 (0.02)	9.0E–28
10: <i>KLRC1</i> , <i>KLRC2</i> , <i>KLRC4</i> , <i>KLRK1</i> , <i>RP11–277P12</i>	P4:3551	NK: %CD314 <sup>–</sup> CD158a <sup>+</sup>	rs2734565	12	C/T	0.3	–0.144 (0.02)	1.34E–10	–0.18 (0.04)	1.67E–05	–0.15 (0.02)	1.4E–14
10: <i>KLRC1</i> , <i>KLRC2</i> , <i>KLRC4</i> , <i>KLRK1</i> , <i>RP11–277P12</i>	P4:4832	NK: %CD314 <sup>–</sup> CCR7 <sup>–</sup>	rs2734565	12	C/T	0.3	–0.233 (0.02)	1.27E–24	–0.33 (0.04)	3.95E–15	–0.26 (0.02)	2.7E–37
10: <i>KLRC1</i> , <i>KLRC2</i> , <i>KLRC4</i> , <i>KLRK1</i> , <i>RP11–277P12</i>	P4:5538	NK: %CD314 <sup>–</sup> CD335 <sup>+</sup>	rs2734565	12	C/T	0.3	–0.275 (0.02)	3.40E–34	–0.38 (0.04)	2.27E–19	–0.30 (0.02)	6.4E–51
11: <i>FTO</i>	P4:3551	NK: %CD314 <sup>–</sup> CD158a <sup>+</sup>	rs1420318	16	A/G	0.1	–0.146 (0.02)	9.34E–11	–0.19 (0.06)	4.05E–02	–0.14 (0.02)	1.2E–11

For the Discovery stage, we used a significance threshold of  $p < 3.3 \times 10^{-10}$ . This threshold corresponds to the standard genome-wide threshold of  $5 \times 10^{-8}$  after further adjustment for 151 independent tests. Orrù et al. (2013) also identified Locus 1 (associated with a single trait, CD62L<sup>–</sup> dendritic cells, not measured in our panel), and Locus 9 (associated with a single trait: CD39<sup>+</sup> CD4 T cell frequency, P2:4186 in our list). The trait ID is fully described in Table S2.



subsets, and CD161 expression on CD4 T cells) were associated with variants on chromosome 1q23 in a genetic region containing the important immune-regulating genes *FCGR2A*, *FCGR2B*, and *FCRLA* (Table 1). These associations were independently verified in the replication cohort, and the combined discovery and replication set p values of the 6 SPELs ranged from  $2.8 \times 10^{-11}$  to  $1.6 \times 10^{-54}$  (Table 1 and Figure 1B). Table S3 illustrates other examples of genetic control of cell-surface expression, including the expression of CD11c, CD123, and CD274 on myeloid subsets.

Overall, 241 SNP variants with a minor allele frequency above 5% were significantly associated with various SPELs (Table S3); of these, 35 SNPs were pleiotropically associated with multiple SPELs.

Genetic control of SPEL may simply be due to promoter/enhancer element variants or more complex regulation of transcription, translation, or protein localization. In contrast, genetic control of CSF may reveal homeostatic mechanisms regulating cell subset representation in the blood. Genome-wide significant associations were identified with 13 different CSFs (Figure 1C and Table 1). Nearly all were verified in the replication cohort (Table 1), and some reached a p value of  $10^{-43}$ .

Suggestive associations, which did not meet the conservative significance threshold of  $3.3 \times 10^{-10}$ , were also identified for numerous SPELs and CSFs (Tables S3 and S4). The associations that were independently replicated (replication  $p < 0.05$ ), as well as meta-analyzed variants reaching  $p < 5 \times 10^{-8}$ , are reported in Table S6.

### Genetic Control of T<sub>REG</sub> Cells

One of the most heritable traits identified from our staining panels was the frequency of CD39<sup>+</sup> cells within the CD4 compartment (Figure 2), as previously reported (Orrù et al., 2013). CD39<sup>+</sup>CD4 T, as well as CD73<sup>+</sup> CD4 T cells, have been identified functionally as T regulatory (T<sub>REG</sub>) cells (Borsellino et al., 2007), a key subset in the modulation of immune responses (Antonioli et al., 2013).

The heritability of CD39<sup>+</sup>CD4 T frequency was 89% (95% CI: 66%–93%) (Figure 2A). GWAS analysis revealed a single locus on chromosome 10 that was highly associated with the trait (Figure 2B); this locus maps to the CD39 gene itself. Quantification of the expression of CD39 on a per-cell basis (i.e., SPEL) revealed that the basis for this association was an “on/off” control of the expression of the CD39 molecule on the cells, rather than a homeostatic regulation of the circulating levels of these cells (i.e., CSF). Specifically, individuals who are homozygous for rs7096317A express the highest amount of this protein on the cell surface; heterozygotes expressed half as much; and rs7096317G homozygotes expressed virtually none (Figures 2C and 2D). Although the A/G heterozygotes have a significantly decreased CD39 SPEL, the cells express enough so as to remain CD39<sup>+</sup>. Thus, in the analysis of the CD39 CSF by genotype, only the G/G homozygotes have a reduced frequency of this population (Figure 2E). This illustrates the power of the SPEL analysis to deconvolute potential mechanisms of genetic control that are missed by simple analysis of CSF.

Similarly, the frequency of CD4 T cells that are CD25<sup>+</sup>CD127<sup>−</sup>CD45RO<sup>+</sup> but do not express CD39 was also strongly associ-

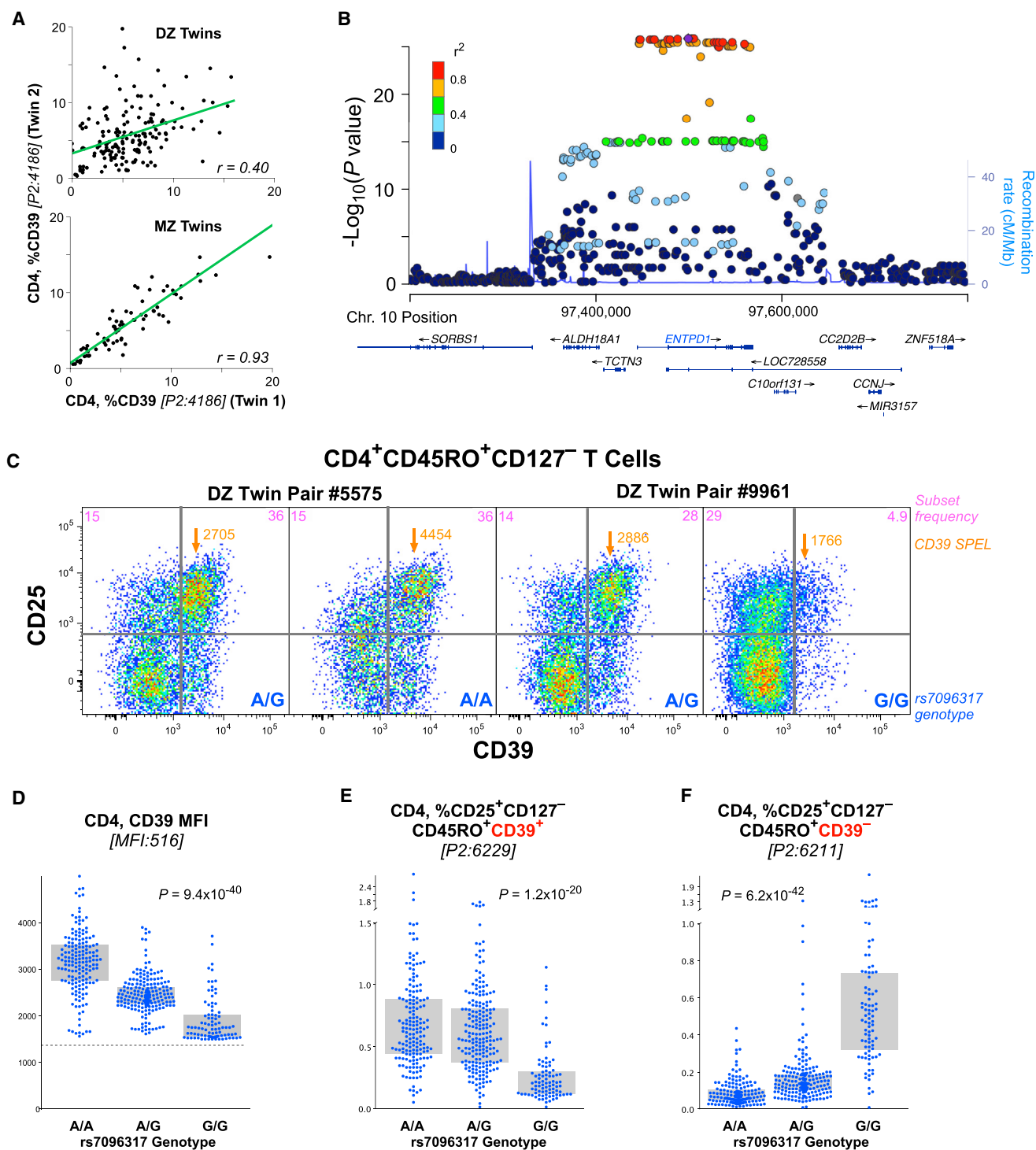
ated with this same locus, showing the opposite association (Figure 2F). In other words, genetic control is not over the frequency of T<sub>REG</sub> (CD4<sup>+</sup>CD25<sup>+</sup>CD127<sup>−</sup>CD45RO<sup>+</sup>) but over the quantitative expression of CD39 (the cell phenotype). Notably, this genetic control also extends to lymphocytes that are not T<sub>REG</sub>: a similar genotypic association (Figure S6) was found for the relatively rare CD8<sup>+</sup> and CD4<sup>−</sup>CD8<sup>+</sup> T cells expressing CD39 (Figures S7A and S7B). Finally, CD73 is an ectonuclease similar to CD39, and its expression has also been associated with T<sub>REG</sub> cells (Antonioli et al., 2013). The expression of CD73 was also found to be genetically controlled (Figure S7C) and associated with a single locus on chromosome 6 mapping to the CD73 gene itself.

Thus, the main genetic control of CD39<sup>+</sup> T<sub>REG</sub> appears to originate from a transcriptional or post-transcriptional regulation leading to the presence or absence of this protein on the cell surface of T<sub>REG</sub> cells; for those T<sub>REG</sub> defined on a basis independent of CD39, we found no evidence of genetic control over their representation in blood.

### Genetic Influences on Leukocyte Differentiation

In virtually every leukocyte population, we found examples in which the frequency of certain differentiation stages was heritable (Figure 3). In some cases, despite a very high heritability, we were unable to identify genetic variants that correlated with the trait. For example, the frequency of a CD4 transitional memory (T<sub>TM</sub>) phenotype (CD28<sup>+</sup>CD127<sup>−</sup>), which comprises 15% to 20% of CD4 T cells, was very strongly heritable (Figure 3A) but did not correlate with any SNP genotypes. We found similar examples of strong heritability without genetic associations for other T cell stages, including recent thymic emigrants (RTE) and central memory (T<sub>CM</sub>). This was not unexpected—our study was only powered to find large effect sizes of gene variants, which is unusual for most traditional disease GWASs that need thousands of subjects per association. This suggests that possibly multiple genes of modest influence act on these phenotypes. Despite the lack of defined genetic association, the observation of strong heritability indicates that these cell types play an important and unique role in immunity, such that their numbers are strongly regulated.

For a number of leukocyte lineages, we were able to identify genetic associations with differentiation stages and illustrate four examples. (1) Within B cells, the proportion that is immature (CD10<sup>+</sup>) is associated with the genotype of the membrane metallo-endopeptidase (*MME*) gene (Figure 3B). (2) The proportion of Th22 CD4 T cells (CXCR3<sup>+</sup>CCR4<sup>+</sup>CCR6<sup>+</sup>CCR10<sup>+</sup>) is strongly associated with a single locus on chromosome 16, mapping to the *SPG7* gene, which codes for paraplegin, an important protein in mitochondrial function (Figure 3C). (3) Within natural killer (NK) cells, the proportion of cells that express CD335 but not CD314 (Figure 3D) maps to the *KLRC4* gene (Figure 3E). This association is much more profound for NK cells that are in an early differentiation stage (CD56<sup>+</sup>CD16<sup>−</sup>) and becomes less strong as the cells mature (Figure 3F). This indicates that a mechanism evinced in a differentiation stage-specific manner. (4) The proportion of T cells that are “stem cell memory” cells (T<sub>SCM</sub>; the earliest memory stage) is heritable (Figure 3G) and associated with a genetic locus containing *FAS* (CD95) (Figure 3H). This



**Figure 2. Genetic Associations with Treg Phenotype Cells**

(A) The correlation of the fraction of CD4 T cells that are CD39<sup>+</sup> in dizygotic twins (top) and monozygotic twins (bottom). The linear correlation,  $r$ , is shown for each comparison.

(B) Locus plot showing significant effect of individual SNPs on CD39 expression on CD4 T cells.

(C) Shown are the expression profiles of CD39 and CD25 for the subset of CD4 T cells that are CD45RO<sup>+</sup>CD127<sup>-</sup> for two pairs of dizygotic twins discordant for the rs7096317 allele (in the CD39 gene locus). Within each graphic is shown the fraction of cells in the upper two quadrants and the surface protein expression level (SPEL) of CD39 for the cells in the upper right quadrant, as well as the genotype of each individual.

(legend continued on next page)

association was much stronger for CD4 than for CD8 T cells.  $T_{SCM}$  are precursor cells that have tremendous proliferative capacity and can regenerate all other memory T cell populations (Gattinoni et al., 2011; Lugli et al., 2013). Interestingly, this same locus also has a significant association with the fraction of T cells that are CD8 (Figure 3I), demonstrating multiple (pleiotropic) effects of the FAS gene on T cell differentiation.

### Pleiotropic Impact of the FcRG2 Locus

The locus with the widest range of impacts on leukocyte subset phenotype and frequency was on chromosome 1, a region including the FcRG2 gene. This locus is well known for its association with a variety of autoimmune and inflammatory diseases, including systemic lupus erythematosus (SLE), Kawasaki's disease, inflammatory bowel disease, Crohn's disease, type 1 diabetes, and HIV disease progression. Despite the genes in this locus being primarily expressed on myeloid and B lymphoid cells, many of these diseases are traditionally associated with T cell dysregulation.

The strongest association (e.g., SNP rs1801274) we identified for this locus was with the expression of CD32 (FcRG2a and/or FcRG2b: these are indistinguishable by the monoclonal antibody used in our panel) on the surface of inflammatory myeloid dendritic cells (imDC; Figures 4A and 4B). The heritability of this trait was extremely high at 96% (CI: 81%–97%).

The genetic control of the expression of CD32 on imDCs was not seen in all cell populations. For example, B cells showed no control (Figure 4C), whereas the expression of CD32 on imDCs is associated with the number of rs1801274 "T" alleles (Figures 4C and 4D).

The rs1801274 genotype has been strongly associated with susceptibility to SLE, as well as another SNP in the same locus, rs10800309. This latter SNP has also been associated with ulcerative colitis. The frequency of CD32<sup>+</sup> imDCs is strongly affected by the genotypes at both of these loci (Figures 4D and 4E); however, the distribution of expression for either locus is not uniform: high, intermediate, and low expressors can be found within all genotypes with differing frequencies. However, when the two genotypes are taken together as a diplotype, a powerful and replicated association becomes evident for CD32<sup>+</sup> imDCs (Figures 4E and S8C). The impact of this diplotype on CD32 expression extends to other myeloid subsets (Figures 4F and S8A). Statistical significance of the association is greatest for monocytes, although the dynamic range in the expression levels is not as wide as it is for imDCs. Other subsets, such as the professional antigen presenting mDC, show a muted control of expression; CD11c<sup>+</sup>CD123<sup>+</sup> DC, like B cells, show no differential regulation of CD32 expression at all.

Given the profound impact of these genotypes on particular subsets, it raises the possibility that part of the increased susceptibility to associated autoimmune diseases may be a consequence of the altered function of cells like imDCs by virtue of a differential expression of the activating (CD32a) or repressing

(CD32b) proteins that we identify here. This is perhaps driven by a SNP in the promoter/enhancer areas in high linkage disequilibrium to the commonly studied coding SNP rs1801274.

We also found a remarkable range of effects of the FcRG2 locus on a variety of lymphocyte subsets (Figure 5). For example, the proportion of early NK cells that are CD2<sup>+</sup>CD158a<sup>+</sup>CD158b<sup>+</sup> is strongly associated with SNP rs365264 (Figure 5A), located between CD32a and CD16 (Figure 4B). The rs1801274 coding SNP in the locus was associated with phenotypes on both B cells and T cells, including the fraction of memory IgG<sup>+</sup> (Figure 5B) or IgA<sup>+</sup> (Figure S8B) B cells that express CD27, as well as the CD27 expression level on a per cell basis. Interestingly, in this case, the higher surface expression levels of CD27 (SPEL) are associated with lower frequencies of cells that express CD27 (CSF). Thus, in contrast to the example of CD39<sup>+</sup> T<sub>REG</sub> (Figure 2), differential regulation of CD27 protein expression does not account for differential frequency of these cell subsets.

### T Cells, FcRG2, and Autoimmune Disease

Similar to the case for IgG<sup>+</sup> B cells, CD8 T cells also exhibited higher CD27 expression in association with rs1801274T allele (Figure 5C); this was also true for other T cell lineages (Figure S8). Furthermore, a population of CD4 T cells that express CD161 is also strongly associated with this same genotype (Figure 5D). Importantly, CD161<sup>+</sup> T cells are either Th17 or mucosal-associated innate T (MAIT) cells, important for maintenance of mucosal integrity. Thus, we define an impact of specific gene variants on important T cell phenotypes closely related to their activation potential, which may underlie the associations with T-cell-based autoimmune diseases.

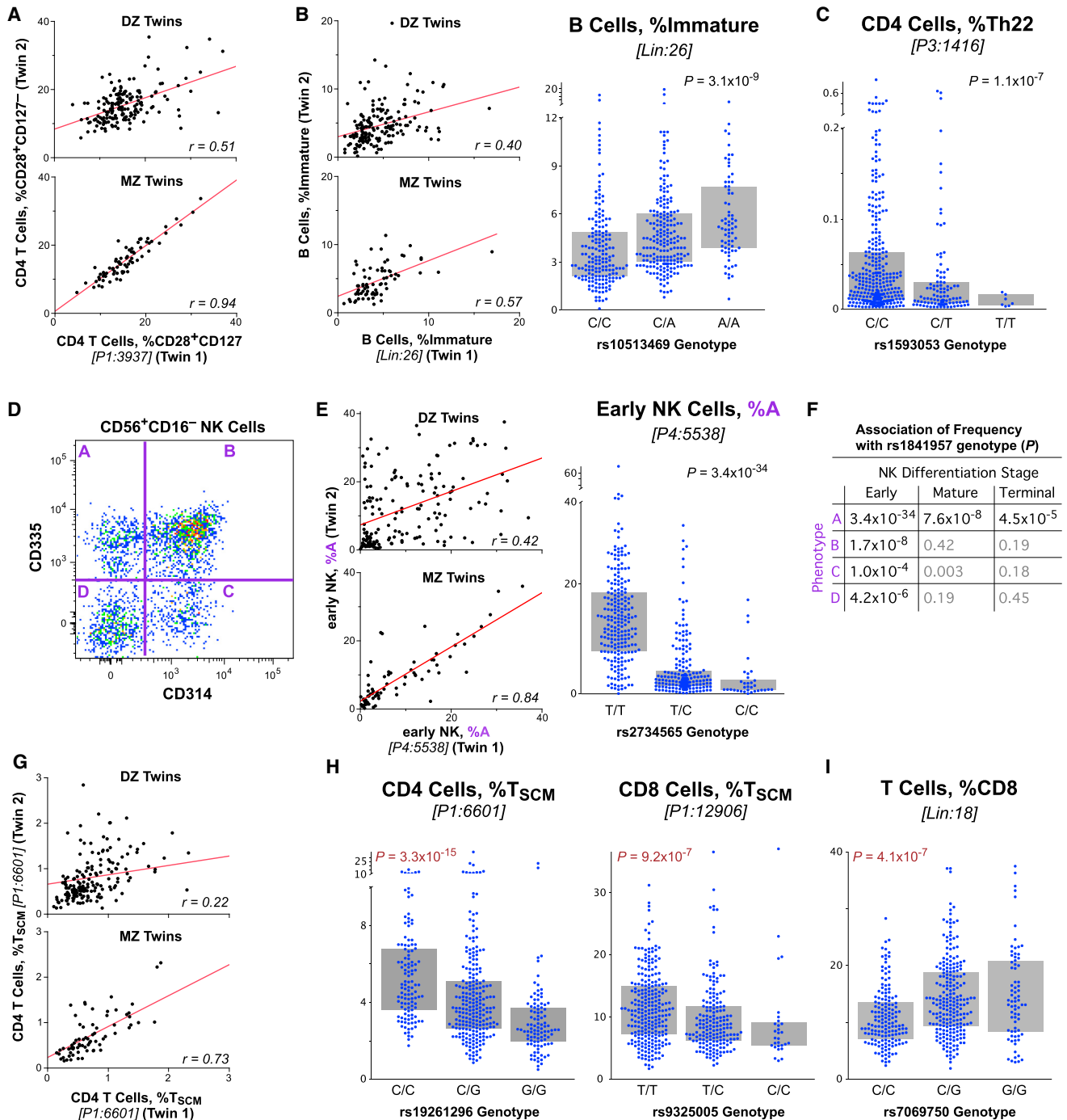
Finally, Starbeck-Miller et al. (2014) recently demonstrated that CD8 T cells can express CD32b and that this expression was functionally important in modulating cytolytic T cell responses. Here, we demonstrate that expression of CD32 on CD8 T cells is low and variable between individuals (Figure 5E). Notably, this expression shows a very strong association with the rs1801274:rs10800309 diplotype of the FcRG2 gene locus (Figure 5F). This suggests that the regulation of surface expression of this negative regulatory molecule on CD8 T cells (Starbeck-Miller et al., 2014) has a common expression mechanism to that in myeloid populations. Gene variants increasing susceptibility to SLE also associate with lower levels of this negative regulatory protein on CD8 T cells and imDC. Together, these data provide a possible direct link between the SNPs highly associated with autoimmune diseases and T cell phenotypes that might account for the pathogenesis.

### Overlap with Disease Associations

The Catalog of Published Genome-wide Association Studies (<http://www.genome.gov/gwastudies/>) and ImmunoBase (<http://www.immunobase.org/>) were used to evaluate the overlap between genetic variants associated with CSFs and SPELs in our study and those reported to be suggestively or statistically

(D) The CD39 SPEL of CD39 positive cells is graphed by the genotype of rs7096317; the dotted line indicates the threshold of positivity above which a cell was considered CD39<sup>+</sup>. In the C/C genotype, relatively few cells are above this threshold and the median fluorescence intensity values are not robust.

(E and F) The fraction of CD4 T cells of the designated phenotype is graphed by the rs7096317 genotype. Bars indicate interquartile range.



**Figure 3. Genetic Associations with Lymphocyte Differentiation**

(A) The proportion of CD4 T cells that are “translational memory” (CD28<sup>+</sup>CD127<sup>-</sup>) is shown for DZ and MZ twins.  
(B) The proportion of B cells that are immature is shown for twins (left) and is strongly associated with the genotype of rs10513469 (MME gene) (right).  
(C) The proportion of CD4 T cells that are Th22 (CXCR3<sup>+</sup>CCR4<sup>+</sup>CCR6<sup>+</sup>CCR10<sup>+</sup>) is associated with the genotype of rs2019604 (SPG7 gene).  
(D) A frequency of four phenotypes within NK cells (designated as “A”...“D” based on the expression of CD314 (KLRC4) and CD335) is shown for “early” (CD56<sup>+</sup>CD16<sup>-</sup>) differentiated NK cells.  
(E) The proportion of early NK cells that are CD314<sup>-</sup>CD335<sup>+</sup> (population “A”) is shown for DZ and MZ twins (left). (Right) The genotypes of rs1841957 (near the KLRC4/CD314 locus) strongly associate with the frequency of CD314<sup>-</sup>CD335<sup>+</sup> cells among early NK.  
(F) The associations of rs1841957 with all four phenotypes within differentiation stages of NK cells are shown by p value.

(legend continued on next page)



significant in candidate SNP, candidate gene, or genome-wide association studies of complex and infectious diseases.

SNPs that were highly correlated variants ( $r^2 > 0.8$ ) with our significant immune traits were also interrogated for overlap with reported disease associations using the appropriate thresholds for the number of tests. A number of gene variants significantly and suggestively ( $p < 10^{-3}$ ) correlated with CSFs and SPELs have been reported in associations of complex and infectious diseases, as shown in Table 2. The different gene variants of *FCGR2A*, associated with a range of myeloid and T cell phenotypes in our data, were reported to be associated with increased risk of a number of diseases, including inflammatory bowel disease, ulcerative colitis, SLE, Kawasaki disease, ankylosing-spondylitis, and HIV progression (Table 2). An additional variant of *FCGR2A*, rs10494359 (associated with P7:100 [CD64<sup>+</sup>CD274<sup>+</sup>imDCs]), is closely correlated (in linkage disequilibrium) with rs10494360 ( $r^2 = 0.941$ ) and has been associated with rheumatoid arthritis. Juvenile idiopathic arthritis and chronic lymphocytic leukemia susceptibility loci in the *ACTA2/FAS* region of chromosome 10q23.31 were also associated with the frequency of P1:6601 (CD4 T<sub>SCM</sub>) ( $p = 4.1 \times 10^{-12}$ ; Table 2). The Behçet's disease susceptibility variant in the killer cell lectin-like receptors gene region corresponded with the frequencies of P4:3551, P4:4832, and P4:5538 (all three are CD314<sup>+</sup> subsets of NK cells). The correlations of variants association with tuberculosis, malaria, leprosy, HIV, and hepatitis B and C our immune traits are presented in Table 2.

## DISCUSSION

Understanding the fundamental principles of how the immune system protects the host from infection yet also contributes to autoimmunity and other disease pathogenesis is essential for the development of novel diagnostics and medicines. There remains a major gap in our understanding of genetic determinants of a normal human immune system and its main coordinates such as the frequency of immune cells and expression of relevant proteins. Using 669 twins and the richest immunophenotyping performed to date, we investigated the genetic architecture of immune traits. We describe multiple independent genetic variants at several genetic loci explaining a substantial proportion (up to 96%) of the genetic variation. We identify both pleiotropic genetic loci that control multiple immune traits and single immune traits under genetic influence by multiple loci. For certain canonical immune traits, genetic control is exerted at the level of immune cell-surface protein expression (i.e., a consequence of promoter/enhancer or signaling mechanisms) rather than at the level of cell subset frequency (i.e., homeostasis or differentiation mechanisms). We further describe multiple genetic associations with common canonical immune traits related to leukocyte lineage and differentiation of major immune cell subsets such as B cells, T cells, and natural killer cells. Finally, we identify genetic elements associated with both immune traits and auto-

immune and infectious diseases. Providing the heritabilities of thousands of cell subtypes plus a basis to uncover the genetic architecture of the numerous gene-immune associations establishes this data set as an essential bioresource for researchers. The remarkably strong associations we find for genetic traits linked to disease illustrate the power of our approach by using twins and optimized high-quality immune phenotyping.

Some limitations of our study should be noted: the cohort used is all female, it is (for GWAS) a relatively limited sample size, and it is relatively homogeneous in terms of environmental exposure. The low numbers of genetic associations on chromosome 6 (the major histocompatibility region) is possibly explained by the considerable complexity and polymorphism in this gene region, which would require larger sample sizes to obtain statistically significant genetic associations. With regard to immunological traits, it should be noted that our discovery cohort ranges in age from 41 to 77. It is possible that analysis of a younger cohort, for which less environmental pressure on the immune system has occurred, would reveal stronger associations; on the other hand, it is likely that the greatest power to detect immune correlates related to disease will come from measurements at a time most proximal to the typical onset. Nevertheless, our success in identifying a large number of genetic variants with genome-wide significance validates our approach of focusing on well-defined and curated immune phenotypes. It should be noted that the 297 SNPs we report (Table S5) are those that attained genome-wide significance (with a conservative correction for multiple comparisons) in both the discovery and replication cohorts. Many more associations are evident in the data set (e.g., Tables S2 and S3), which can serve to formulate new testable hypotheses and genetic studies.

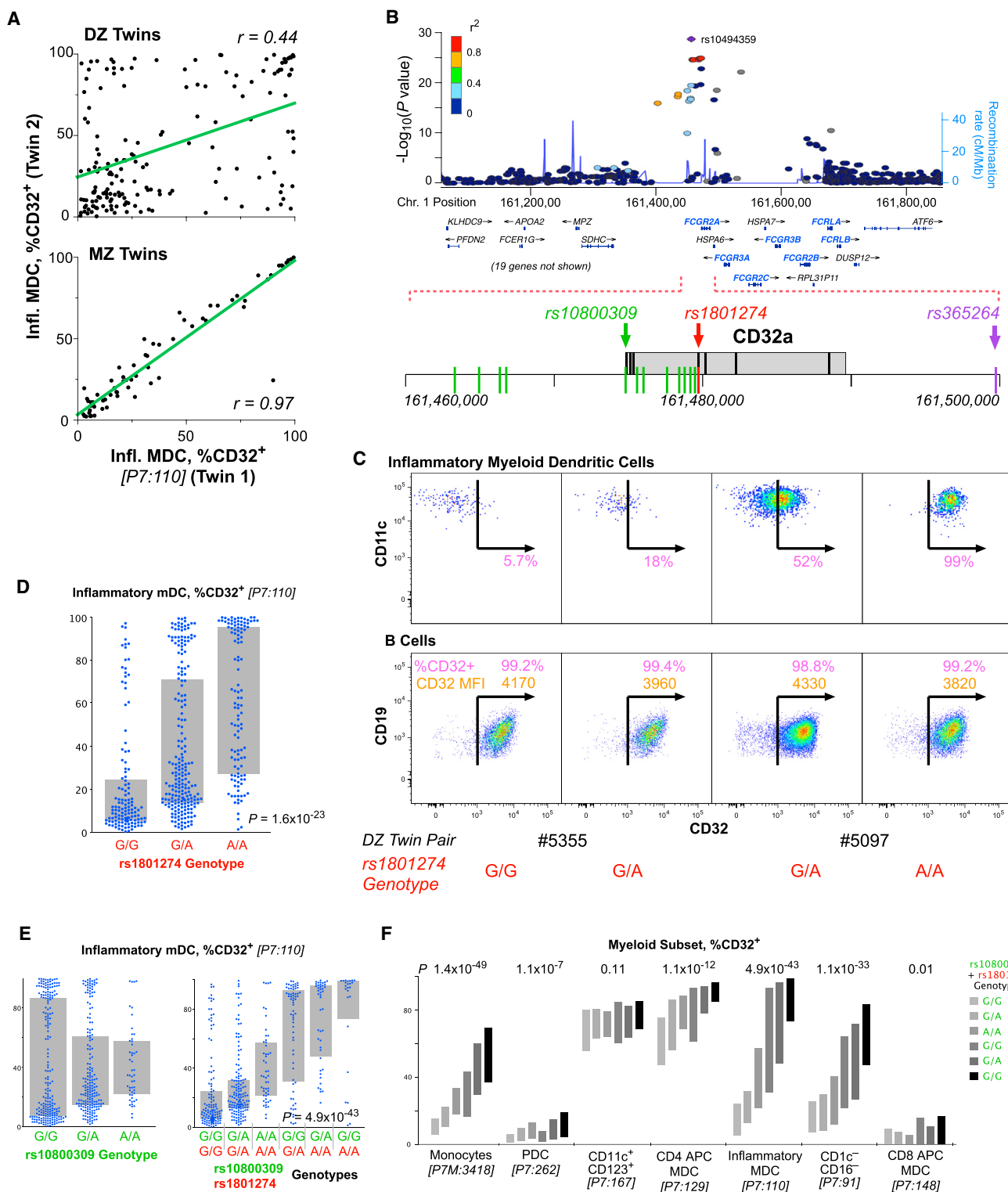
An example of the power of the resource was in distinguishing two important mechanisms that lead to differential representation of immune cell phenotypes (e.g., CD39<sup>+</sup> CD4 T cells). The first is homeostatic—i.e., mechanisms that control the recirculation, proliferation, and elimination of a certain cell type in the blood—a cellular mechanism expressed at the whole-body level. Such mechanisms, although unidentified, are known to exist for regulating major subset numbers (such as CD4 T cell numbers). A second, completely independent mechanism is the molecular regulation of the protein expression at the cell level itself (e.g., promoter/enhancer variants). Thus, even in the presence of intact homeostatic mechanisms regulating that cell type, a reduction in the number of cells expressing a protein (perhaps part of the defining phenotype) may be due to a specific promoter variant that simply abrogates the expression of the gene. Here, we show examples of both mechanisms.

An example of a major immune trait under genetic control is the phenotype, but not frequency, of regulatory T (T<sub>REG</sub>) cells. T<sub>REG</sub> cells are essential for maintenance of immune homeostasis, and their dysregulation might lead to autoimmunity (Sakaguchi et al., 2010). A pleiotropic locus containing the ectonucleoside triphosphate diphosphohydrolase 1 (*ENTPD1* and CD39)

(G) The proportion of CD4 T cells that are "stem cell memory" (T<sub>SCM</sub>: CD45RA<sup>+</sup>CD95<sup>+</sup>CD27<sup>+</sup>CD28<sup>+</sup>CD127<sup>+</sup>CD57<sup>+</sup>) is shown for DZ and MZ twins.

(H and I) The genotypes of rs7069750 (*FAS* gene) are associated with the proportion of CD4 and CD8 T cells that are T<sub>SCM</sub>, as well as the proportion of all T cells that are CD8.

Bars indicate interquartile ranges.



**Figure 4. Genetic Associations of the FcR Locus with Myeloid Immunophenotypes**

(A) The correlation of the fraction of imDCs that are CD32<sup>+</sup> for dizygotic twins (top) and monozygotic twins (bottom). The linear correlation,  $r$ , is shown for each comparison.

(legend continued on next page)

gene controls several phenotypic features of CD39<sup>+</sup> T<sub>REG</sub>. Although we confirm an apparent association of this gene locus with the frequency of circulating CD39<sup>+</sup> T<sub>REG</sub> (Orrù et al., 2013), we show here that this is a consequence solely of altered phenotype. Quantitative analysis of CD39 protein expression on the cell surface demonstrates that the genetics of control of CD39<sup>+</sup> T<sub>REG</sub> is exerted at the level of surface protein expression, rather than cell frequency (homeostasis). This establishes a paradigm for future studies of immune traits and points to the necessity of including both cell frequency and cell protein expression analyses in immunogenetic studies of the human immune system. We also describe the association of a genetic locus containing the ecto-5'-nucleotidase (*NT5E* and *CD73*) gene with a population of CD73<sup>+</sup> T<sub>REG</sub>. From a functional perspective, it is of interest that both *CD73* and *CD39* are ectonucleotidases involved in the generation of immunosuppressive adenosine that alters T cell and NK cell activities (Deaglio et al., 2007). Thus, it appears that the *CD39* and *CD73* adenosine immunosuppressive pathway has been under evolutionary selection and might therefore be a critical determinant of functional T<sub>REG</sub> activity (Bastid et al., 2013) and establishment of tissue homeostasis. Importantly, we conclude that there is no genetic control of the frequency of T<sub>REG</sub> (i.e., CD127<sup>+</sup>CD25<sup>+</sup> memory CD4 cells), but rather, control is evinced at the level of their specific phenotype and presumably function. Furthermore, the lack of a known disease association with this locus calls into question the importance of *CD39* expression for T<sub>REG</sub> function.

We discovered several genetic associations with immune traits relevant to lymphocyte lineage and differentiation. A genetic locus containing the cell-surface death receptor *FAS* (*CD95*) was associated with the frequency of circulating T stem cell memory cells (T<sub>SCM</sub>). T<sub>SCM</sub> are a recently described infrequent and functionally important lymphocyte subset (Gattinoni et al., 2011); these cells have a largely naive T cell phenotype but are able to self-renew while displaying functional attributes of memory cells. Genetic control at the level of *CD95* suggests a potential role of *CD95* in the control of T stem cell homeostasis for not only differentiation stages such as T<sub>SCM</sub> but total CD8 as well (Figure 3). It further provides a possible link between human autoimmune syndromes based on genetic *CD95* deficiency (Strasser et al., 2009) and T<sub>SCM</sub>.

A genetic locus within a cluster of genes referred to as the "NK complex" containing *NKG2D* (*CD314* and *KLRK1*) was associated with the frequency of a distinct population of CD314<sup>+</sup>CD335<sup>+</sup> "early" (CD56<sup>+</sup>CD16<sup>+</sup>) NK cells. The *NKG2D* gene encodes for a C-type lectin protein preferentially expressed in NK cells. It binds to a diverse family of stress-induced ligands

that include MHC class I chain-related A and B proteins (*MICA* and *MICB*), essential for the activation of T cells and NK cells (Raulet et al., 2013). These data establish an unexpected link between the genetic control of the frequency of a specific subset of NK cells and a gene locus containing major genes with functional relevance to NK cell activity and their activation by stressed normal tissue cells or tumor cells.

One locus in chromosome 1q23 containing *FCGR2A*, *FCGR2B*, and *FCRLA* was associated with multiple immune traits. FcGR genes encode immunoglobulin Fc surface receptors found on macrophages, dendritic cells, and neutrophils, as well as B and NK cells, and are involved in the regulation of B cell antibody production and phagocytosis of immune complexes. The main associated immune traits were the frequencies of CD32<sup>+</sup> inflammatory dendritic cells and monocytes, as well as several T cell phenotypes. Genetic variation in the FcGR gene locus is associated with an increase in susceptibility to several autoimmune and infectious disease, including SLE, ankylosing spondylitis, HIV progression, and several other syndromes. Given the profound impact of this gene locus on particular immune cell subsets, altered function of CD32<sup>+</sup> dendritic cells could be key to increased susceptibility to these autoimmune and infectious diseases.

Indeed, our demonstration of an association between the *FCGR2A* SNP and T cell phenotypes and/or inflammatory myeloid cells (e.g., imDCs) provides a potential link between this locus and autoimmune diseases with T cell etiology. This coding SNP results in variants of the Fc receptor that have different avidities for immunoglobulin and C-reactive protein; consequently, much current research is aimed at understanding the possible functional role of these alleles in autoimmunity. However, our data suggest a different possibility with a more proximal mechanistic link: that association is with a promoter/enhancer SNP (in strong linkage disequilibrium with the coding *FCGR2A* SNP) that modulates expression of the negative regulatory CD32b molecule on imDC, monocytes, and/or T cells.

In addition to the wide range of diseases associated with the FcRG locus, further examples include a SNP within a genetic locus associated with Behçet's disease (Kirino et al., 2013) that is in tight linkage disequilibrium with a SNP controlling the frequency of CD314<sup>+</sup>CD335<sup>+</sup> early NK cells. We also report that a genetic locus containing *FAS* and associated with Juvenile idiopathic arthritis (Hinks et al., 2013) is also associated with functionally important CD8<sup>+</sup> T stem memory cells.

These findings illustrate a key value of our database and approach: the identification of candidate immune traits associated with genetic loci of relevance to autoimmunity and infection. In summary, using one of the most comprehensive immune

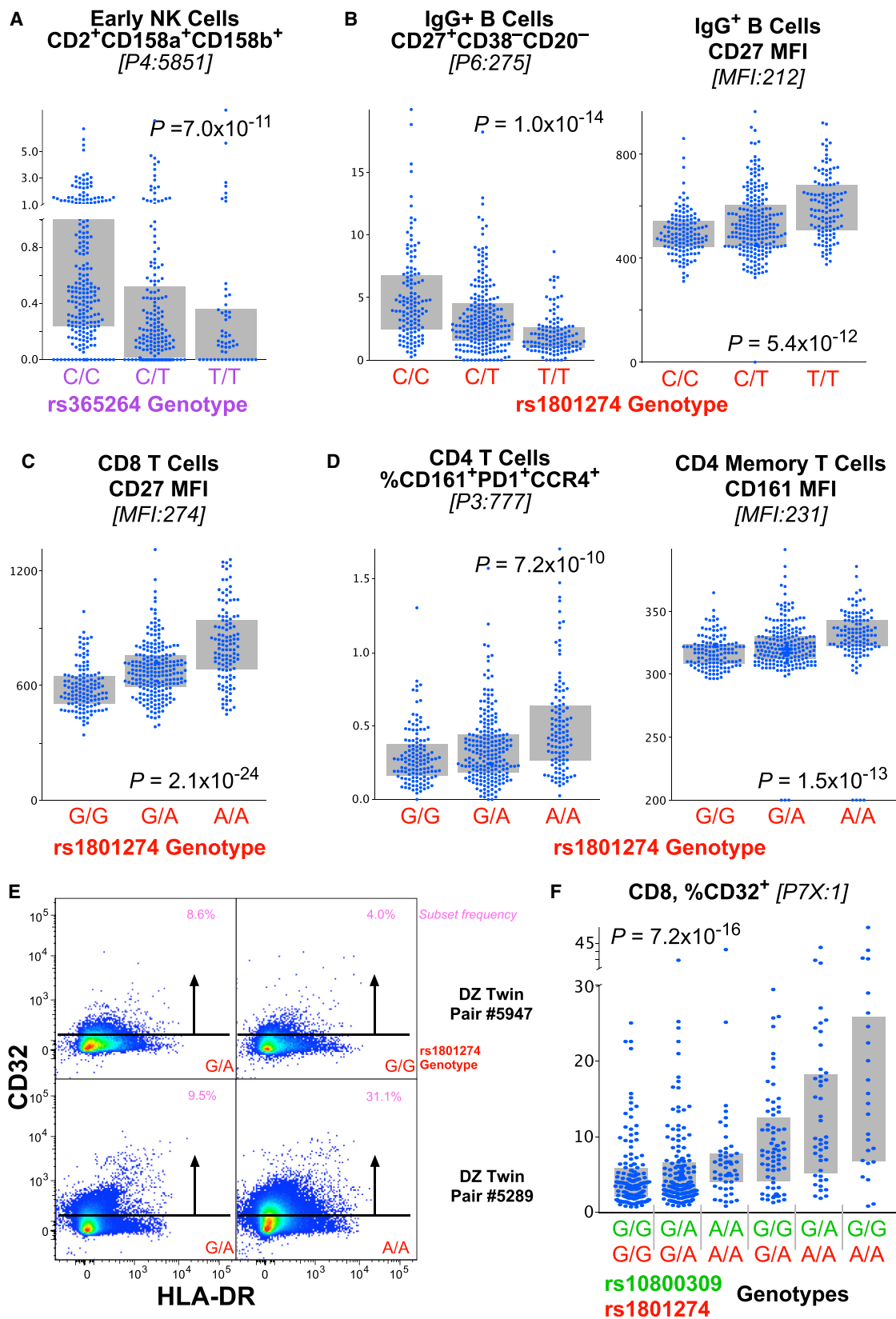
(B) Organization of the FcR locus of chromosome 1 showing the position of immunologically relevant genes. The positions of three SNPs are highlighted in color; rs1801274 and rs10800309 are the two that are most closely associated to susceptibility to SLE. SNPs shown in green were in complete linkage disequilibrium within the samples analyzed in our cohort.

(C) Sample expression profiles of CD32 on imDCs (top) and B cells (bottom). Shown are the fraction of cells that are CD32<sup>+</sup> (in pink) and, for the B cells, the CD32 SPEL (in orange). Two pairs of dizygotic twins discordant for the genotype at rs1801274 are shown.

(D) The distribution of expression of CD32<sup>+</sup> imDCs is shown by genotype at rs1801274.

(E) The expression of CD32 on the imDCs is not significantly associated with the genotype at rs10800309 by standard ANOVA ( $p = ns$ ); however, the distributions are clearly different by genotype. The combination of the genotype at rs10800309 and rs1801274 provides a dramatic distinction for the expression of CD32 on imDC.

(F) The expression profile of CD32 on seven different myeloid populations is shown, broken down by the combined genotype of the two SNPs. Bars indicate interquartile range.



(legend on next page)



phenotyping efforts to date, we identified numerous genetic loci controlling key parameters of a normal human immune system. This comprehensive human immune phenotyping bioresource will allow the identification of critical immune phenotypes associated with common autoimmune and infectious diseases, ultimately leading to accelerated discovery of mechanisms of disease and response to therapy.

## EXPERIMENTAL PROCEDURES

### Samples

This study was approved by the NIAID (NIH) IRB and London-Westminster NHS Research Ethics Committee; all participants provided informed consent. The discovery stage comprised 497 female participants from the UK Adult Twin Register, TwinsUK, with full genotyping data on 460 subjects. The TwinsUK cohort is described in detail in (Moayyeri et al., 2012). Briefly, TwinsUK is a large cohort of twins historically developed to study the heritability and genetics of diseases with a higher prevalence among women. The study population is not enriched for any particular disease or trait and is representative of the British general population of Caucasian ethnicity. Selected twins were all female with an age range of 41–77 (mean 61.4), and by self-report, 100% Caucasian with most being UK ancestry. From the subsequent genotype data, we excluded a few individuals showing evidence of non-European ancestry as assessed by principal component analysis comparison with HapMap3.

The replication stage included a further unrelated 172 TwinsUK participants with whole-genome genotyping data on 169. The samples for the discovery samples were selected to match the characteristics (age and gender) of the discovery data set. For this reason, the replication cohort included only Caucasian women with an age range of 32–83 (mean 58.2). All subjects were nominally healthy at the time of sample collection.

A total of 746 PBMC vials were analyzed: PBMC from the 669 twin specimens (plus 4 replicates) and 30 healthy controls from the US (two vials of PBMC from blood drawn six months apart were analyzed for 29 subjects, a replicate vial of PBMC from one of the two blood draws for 14 (of the 29), and one vial only for 1 subject). The samples in each stage were ordered such that twin or longitudinal control samples were analyzed in the same experimental run (each comprising 15–30 vials), whereas replicate control samples were analyzed in different experimental runs. Staining and data analyses were otherwise performed blinded to identity.

### Immunophenotyping

See [Extended Experimental Procedures](#) for cell processing details.

### Flow Cytometry and Data Analysis

Cells were analyzed in 96-well plate format on an 18-color LSR (BD Biosciences) using an HTS unit. Each run on the flow cytometer was accompanied by a set of compensation controls of antibody-stained IgG kappa beads (BD Biosciences). Data were evaluated on FlowJo software v9.7 (FlowJo, LLC). Post-processing of data and visualizations were done with JMP v10 (SAS) and SPICE v5.3 (NIAID; Roederer et al., 2011).

### Gating

A graphic depicting the fluorescence distribution of all samples in the discovery cohort is shown in [Figure S1](#). [Figures S2](#) and [S3](#) illustrate the gating hierarchy for a single sample. The [Extended Experimental Procedures](#) describe the generation of the ~80,000 gates analyzed.

For the discovery cohort, there were 20 experimental runs for 543 samples (501 twin specimens from 497 subjects and 42 US control specimens from 14 subjects). Within each run, uniform scatter gating was used. Each sample was gated on time (to eliminate spurious events from beginning or end of sample run); this gate could vary by sample. With two major exceptions, all samples received the same fluorescence gating. After the third run, we chose to replace the CD4 reagent in panels 1, 2, and 5 due to poor performance; this necessitated a different CD4 gating for those samples. Similarly, the reagents in the “dump” channel of panel 7 were modified after 6 runs.

For the replication cohort, there were 8 experimental runs for 203 samples (172 twin specimens from 172 subjects and 31 US control specimens from 16 subjects). All analysis procedures were identical to the discovery cohort. Minor modifications to the panels were necessitated by unavailability of the same lots of reagents, but these did not impact enumeration of subsets. Specimen processing for the replication cohort was initiated after final analysis of the discovery cohort.

### Genotyping

Genotyping was conducted with a combination of Illumina arrays (HumanHap300 and HumanHap610Q) (Richards et al., 2008; Soranzo et al., 2009). The Illuminus calling algorithm (Teo et al., 2007) was used to assign genotypes. No calls were assigned if an individual's most likely genotype was called with less than a posterior probability threshold of 0.95. Validation of pooling was achieved via a visual inspection of 100 random, shared SNPs for overt batch effects. Finally, intensity cluster plots of significant SNPs were visually inspected for over-dispersion, biased no calling, and/or erroneous genotype assignment. SNPs exhibiting any of these characteristics were discarded. Stringent QC measures were performed on the genotypes prior to data analysis. The sample exclusion criteria were: (1) sample call rate < 98%, (2) heterozygosity across all SNPs  $\geq 2$  SD from the sample mean; (3) evidence of non-European ancestry as assessed by principle component analysis comparison with HapMap3 populations; (4) observed pairwise identity by descent (IBD) probabilities suggestive of sample identity errors; (5) misclassified monozygotic and dizygotic twins were corrected based on IBD probabilities. The exclusion criteria for SNPs were: (1) Hardy-Weinberg equilibrium (HWE)  $p$  value <  $10^{-6}$ , assessed in a set of unrelated samples; (2) minor allele frequency (MAF) < 1%, assessed in a set of unrelated samples; (3) SNP call rate < 97% (SNPs with MAF  $\geq 5\%$ ) or < 99% (for  $1\% \leq \text{MAF} < 5\%$ ). Alleles of both data sets from the genotyping arrays were aligned to HapMap2 or HapMap3 forward strand alleles. Imputation was performed using the IMPUTE v2 software package (Howie et al., 2009). After imputation of the 2,986,407 SNPs available for analysis, 1,419,558 SNPs passed further QC (call rate  $\geq 95\%$ , MAF  $\geq 0.05$ , HWE  $\geq 10^{-4}$ ) and were used for analysis.

### Statistical Analyses

#### Selection of Subsets for Heritability and GWAS Analysis

Full details are in the [Extended Experimental Procedures](#). In brief, we eliminated CSFs with frequencies below 0.1% or above 99% ([Figure S4A](#)). From these, we selected ~200 for in-depth analysis based on heritability or description in the literature (“canonical”).

#### Genome-wide Association Analysis

Because of relatedness in the TwinsUK cohort, we utilized the GenABEL software package (Aulchenko et al., 2007), which is designed for GWAS analysis of family-based data by incorporating pairwise kinship matrix calculated using

## Figure 5. Genetic Associations of the FcR Locus with Lymphoid Immunophenotypes

- (A) The genotype of rs365264 (close to CD16a on chromosome 1) is strongly associated CD56<sup>+</sup>CD16<sup>+</sup> (“early”) NK cells that are CD2<sup>+</sup>CD158a<sup>+</sup>CD158b<sup>+</sup>.  
 (B) The genotype of rs1801274 is associated with the frequency of memory IgG<sup>+</sup> B cells that are CD27<sup>+</sup>CD38<sup>+</sup>CD20<sup>+</sup> as well as the fraction of CD27<sup>+</sup> cells. Note that, for this case, a lower frequency of the subset (left) is associated with higher protein expression (right).  
 (C) Similarly, the genotype of rs1801274 is strongly associated with the cell-surface expression of CD27 on CD8 T cells.  
 (D) The genotype of rs1801274 is strongly associated with the cell-surface expression of CD161 on CD4 T cells, as well as CD4 T cells that are CD161<sup>+</sup>PD1<sup>+</sup>CCR4<sup>+</sup>.  
 (E) CD8 T cells express low levels of CD32 depending on genotype as shown by flow cytometry.  
 (F) The fraction of CD8 T cells that express CD32 is strongly associated with the rs10800309:rs1801274 diplotype (see [Figure 4](#)).

Bars indicate interquartile ranges.

**Table 2. Overlapping Associations with Complex Diseases**

Immune Trait Association						Disease Association			
Gene (Chr)	Marker	Trait ID	Trait Phenotype	Beta	p Value	Reported SNP	Disease	Best P	Reference
FCGR2A (1)	rs1801274	P7:110	iMDC: %CD32 <sup>+</sup>	0.2	1.6E−23	rs1801274 (A)	IBD	2.1E−38	(Jostins et al., 2012)
		MFI:189	CD27 on IgA <sup>+</sup> B	0.13	6.5E−11		Kawasaki disease	7.4E−11	(Khor et al., 2011)
		MFI:212	CD27 on IgG <sup>+</sup> B	0.14	5.4E−12		ulcerative colitis	2.2E−20	(Anderson et al., 2011)
		MFI:231	CD161 on CD4 T	0.13	2.6E−11		ankylosing-spondylitis	1.4E−09	(Cortes et al., 2013)
		MFI:504	CD27 on CD4 T	0.15	5.4E−14		SLE	6.8E−07	(Harley et al., 2008)
		MFI:552	CD27 on CD8 T	0.19	1.3E−21		HIV progression	1.0E−04*	(Forthal et al., 2007)
		P7:224	CD1c <sup>+</sup> mDC: %CD32	0.12	4.4E−09		lymphoma	0.006*	(Wang et al., 2006)
								malaria	0.013*
FCGR2A (1)	rs10494359	P7:110	iMDC: %CD32 <sup>+</sup>	0.34	2.5E−29	rs10494360 (r <sup>2</sup> = 0.94)	rheumatoid arthritis	9.3E−05*	(Eyre et al., 2012)
ZNF804A (2)	rs6755404	P2:3367	Effector CD4: CD127 <sup>−</sup> PD1 <sup>−</sup>	0.09	3.8E−04	rs6755404 (A)	malaria	1.2E−06	(Band et al., 2013)
MIR216B (2)	rs6751715	P3:5372	CD8 T: %CXCR3 <sup>−</sup> R4 <sup>+</sup> R6 <sup>+</sup> R10 <sup>−</sup>	−0.09	1.5E−05	rs6751715	HIV	1.1E−06	(Fellay et al., 2009)
		P3:5661	CD8 T: %CD161 <sup>+</sup>	−0.07	5.1E−04				
		P6:197	IgA+ B: %CD27 <sup>−</sup> CD38 <sup>−</sup>	0.07	7.0E−04				
		P3:5658	CD8 T: %CD161 <sup>+</sup> PD1 <sup>−</sup>	−0.07	7.8E−04				
LOC100505836 (3)	rs2593321	Lin:19	%NKT	0.08	5.5E−04	rs2593321	HIV-1 control	7.7E−06	(Pelak et al., 2010)
MICA (6)	rs4418214	P6:112	IgE <sup>+</sup> B: %CD20 <sup>−</sup> CD27 <sup>−</sup> CD38 <sup>+</sup>	0.12	7.6E−04	rs4418214 (C)	HIV-1 control	1.4E−34	(Pereyra et al., 2010)
BTNL2 (6)	rs3817963	P2:12609	CD8 T: %CD25 <sup>+</sup> CD38 <sup>+</sup> 45RO <sup>+</sup>	−0.08	3.5E−04	rs3817963 (A)	hepatitis C liver cirrhosis	1.3E−08	(Urabe et al., 2013)
		P2:10486	CD8 T: %CD25 <sup>+</sup> CD38 <sup>+</sup>	−0.08	4.1E−04				
HLA-DQB1 (6)	rs2856718	P2:12609	CD8 T: %CD25 <sup>+</sup> CD38 <sup>+</sup> 45RO <sup>+</sup>	−0.08	5.3E−04	rs2856718 (A)	hepatitis B	4.0E−37	(Mbarek et al., 2011)
		P2:10486	CD8 T: %CD25 <sup>+</sup> CD38 <sup>+</sup>	−0.07	7.8E−04				
EHMT2 (6)	rs652888	MFI:578	CD3 on CD8 T	−0.08	5.7E−04	rs652888	chronic hepatitis B	7.1E−13	(Kim et al., 2013)
MTCO3P1 - HLA-DQA2 (62)	rs4273729	P2:10486	CD8 T: %CD25 <sup>+</sup> CD38 <sup>+</sup>	0.08	1.4E−04	rs4273729	chronic hepatitis C	1.7E−16	(Duggal et al., 2013)
		P2:12609	CD8 T: %CD25 <sup>+</sup> CD38 <sup>+</sup> 45RO <sup>+</sup>	0.08	2.3E−04				
ADGB (6)	rs2275606	P1:12906	CD8 T: %TSCM	0.13	6.3E−04	rs2275606 (A)	leprosy	3.9E−14	(Zhang et al., 2011)
ABO (9)	rs8176722	MFI:1	CD123 on mDC	−0.14	1.4E−04	rs8176722 (A)	malaria	8.9E−10	(Band et al., 2013)
ACTA2, FAS (10)	rs7069750	P1:6601	CD8 T: %TSCM	−0.14	4.1E−12	rs7069750 (C)	juvenile idiopathic arthritis	2.9E−08	(Hinks et al., 2013)
						rs2147420 (r <sup>2</sup> = 1)	CLL	3.1E−13	(Berndt et al., 2013)
KLRC4-KLRK1 (12)	rs1049172	P4:3551	NK: %CD314 <sup>−</sup> CD158a <sup>+</sup>	−0.13	8.8E−09	rs2617170 (r <sup>2</sup> = 0.922)	Behçet’s disease	1.3E−09	(Kirino et al., 2013)
		P4:4832	NK: %CD314 <sup>−</sup> CCR7 <sup>−</sup>	−0.23	7.6E−22				
		P4:5538	NK: %CD314 <sup>−</sup> CD335 <sup>+</sup>	−0.27	1.4E−30				
MMP16 (8)	rs160441	P4:3551	NK: %CD314 <sup>−</sup> CD158a <sup>+</sup>	−0.08	2.3E−05	rs160441	tuberculosis	8.4E−06	(Thye et al., 2010)
ARHGAP20 (11)	rs1469170	P4:3551	NK: %CD314 <sup>−</sup> CD158a <sup>+</sup>	−0.08	5.6E−04	rs1469170 (A)	malaria	8.0E−08	(Band et al., 2013)

Association results from the discovery cohort immune trait analyses are reported in the first six columns. The trait ID is fully described in Table S2. The disease-associated variant, pathology, the disease-associated SNP's best-reported p value are indicated in the 7<sup>th</sup>–9<sup>th</sup> columns, respectively. The risk alleles presented correlate with increased disease susceptibility and the r<sup>2</sup> between disease SNP and immune trait, if different, are reported in parentheses after the reported SNP. \*Disease association is significant with candidate SNP method/gene approach.

genotyping data in the polygenic model to correct relatedness and hidden population stratification. The score test implemented in the software was used to test the association between a given SNP and the trait. Additional quality control was conducted on the association results: minor allele frequency (MAF) > 0.1, Hardy-Weinberg equilibrium <  $10^{-8}$ , SNP call rate > 90%. For our results, we used a genome-wide significance threshold of  $p < 3.3 \times 10^{-10}$ . This threshold represents a standard genome-wide significance threshold ( $p < 5 \times 10^{-8}$ ) further adjusted for 151 independent tests (the number of traits analyzed for GWAS in the discovery cohort). Because there is a high intercorrelation among some of the 151 analyzed traits, this is a very conservative approach and ensures the robustness of our findings. Genome-wide validations were also performed on normalized residuals (after corrections for age), with GenABEL taking into account family structure in the model. The validation  $p$  value threshold was set to  $p < 0.05$ .

Results from the genome-wide analyses of all the analyzed traits for both the discovery and the replication cohorts were meta-analyzed. Fixed effects inverse-variance weighted meta-analyses were conducted using METAL (Willer et al., 2010), with significance evaluated at  $p < 5 \times 10^{-8}$ .

Finally, correlation against the FCRG “diplotype” shown in Figures 4 and S8 was done using JMP; ANOVA  $p$  values without correction for family structure are reported.

#### Identifying Traits Correlated with SNPs

Once we had identified significant associations between the selected 151 traits and all SNPs based on standard analyses, we looked for pleiotropic effects of those SNPs. Approximately 1,200 SNPs from about 18 unlinked loci were correlated against all traits.

To correct for multiple comparisons, we evaluated the covariance among all traits. As shown in Figure S4B, there are less than 1,800 traits that show a covariance less than 0.7. Thus, Bonferroni correction sets a significance threshold of  $p < 2.8 \times 10^{-5}$ . In our analyses, which covered about 20 unlinked loci, we used a more conservative threshold, requiring a Wilcoxon rank association of trait values with SNP genotype to be significant if  $p < 10^{-7}$ .

#### GWAS Catalog Look-Ups

In order to ascertain whether the variants from the associations of the immune cell subsets analyses overlapped with significant disease associations, the GWAS catalog (<http://www.genome.gov/gwastudies/>), Immunobase (<http://immunobase.org/page/Welcomes/display>), and SNPedia (<http://www.snpedia.com>) repositories were used. Gene regions and variants reaching genome-wide significance were searched on these repositories and overlapping associations or correlations with proxy SNPs or those in strong LD are reported in Table 2.

#### BioData Repository

Raw and summary data are available for downloading and analysis. Genotype data are available upon request to the authors. See Extended Experimental Procedures for detailed downloading instructions.

#### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, eight figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.046>.

#### AUTHOR CONTRIBUTIONS

M.R., T.D.S., and F.O.N. designed and supervised the study; I.T., L.N., M.T.B., and F.V. processed samples, which were coordinated by P.D.M.; M.H.B., Y.M., P.C., and P.D.M. performed experimental work; M.R., L.Q., M.M., and M.H.B. performed primary data analysis; M.R., L.Q., M.M., and C.M. performed statistical analyses; and M.R., L.Q., M.M., P.D.M., T.D.S., and F.O.N. wrote the manuscript.

#### ACKNOWLEDGMENTS

This work was supported by the Vaccine Research Center (NIAID, NIH) intramural research program and by the Department of Health via the National Insti-

tute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St. Thomas' NHS Foundation Trust in partnership with King's College London, King's College Hospital NHS Foundation (guysbrc-2012-1) Trust, and Dunhill Medical Trust. TwinsUK is also supported by the Wellcome Trust, and TDS is an ERC Advanced Researcher. We wish to thank Kaimei Song, Steve Peretto, Richard Nguyen, Lynda Myles, Gabriela Surdulescu, Dylan Hodgkiss, and Ayrin Nessa for technical help with the samples, as well as the TwinsUK volunteers for participation.

Received: October 9, 2014

Revised: December 22, 2014

Accepted: February 4, 2015

Published: March 12, 2015

#### REFERENCES

- Ahmadi, K.R., Hall, M.A., Norman, P., Vaughan, R.W., Snieder, H., Spector, T.D., and Lanchbury, J.S. (2001). Genetic determinism in the relationship between human CD4+ and CD8+ T lymphocyte populations? *Genes Immun.* 2, 381–387.
- Amadori, A., Zamarchi, R., De Silvestro, G., Forza, G., Cavatton, G., Danieli, G.A., Clementi, M., and Chieco-Bianchi, L. (1995). Genetic control of the CD4/CD8 T-cell ratio in humans. *Nat. Med.* 1, 1279–1283.
- Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A., et al. (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat. Genet.* 43, 246–252.
- Antonoli, L., Pacher, P., Vizi, E.S., and Haskó, G. (2013). CD39 and CD73 in immunity and inflammation. *Trends Mol. Med.* 19, 355–367.
- Aulchenko, Y.S., Ripke, S., Isaacs, A., and van Duijn, C.M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296.
- Band, G., Le, Q.S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., Sisay-Joof, F., Bojang, K., Pinder, M., Sirugo, G., et al.; Malaria Genomic Epidemiology Network (2013). Imputation-based meta-analysis of severe malaria in three African populations. *PLoS Genet.* 9, e1003509.
- Bastid, J., Cottalorda-Regairaz, A., Alberici, G., Bonnefoy, N., Eliaou, J.F., and Bensussan, A. (2013). ENTPD1/CD39 is a promising therapeutic target in oncology. *Oncogene* 32, 1743–1751.
- Berndt, S.I., Skibola, C.F., Joseph, V., Camp, N.J., Nieters, A., Wang, Z., Cozen, W., Monnereau, A., Wang, S.S., Kelly, R.S., et al. (2013). Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat. Genet.* 45, 868–876.
- Borsellino, G., Kleiweiefeld, M., Di Mitri, D., Sternjak, A., Diamantini, A., Gionetto, R., Hopner, S., Centonze, D., Bernardi, G., Dell'Acqua, M.L., et al. (2007). Expression of ectonucleotidase CD39 by Foxp3+ Treg cells: hydrolysis of extracellular ATP and immune suppression. *Blood* 110, 1225–1232.
- Casanova, J.L., and Abel, L. (2004). The human model: a genetic dissection of immunity to infection in natural conditions. *Nat. Rev. Immunol.* 4, 55–66.
- Clementi, M., Forabosco, P., Amadori, A., Zamarchi, R., De Silvestro, G., Di Gianantonio, E., Chieco-Bianchi, L., and Tenconi, R. (1999). CD4 and CD8 T lymphocyte inheritance. Evidence for major autosomal recessive genes. *Hum. Genet.* 105, 337–342.
- Cortes, A., Hadler, J., Pointon, J.P., Robinson, P.C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., Lee, S., et al.; International Genetics of Ankylosing Spondylitis Consortium (IGAS); Australo-Anglo-American Spondyloarthritis Consortium (TASC); Groupe Française d'Etude Génétique des Spondylarthrites (GFECS); Nord-Trøndelag Health Study (HUNT); Spondyloarthritis Research Consortium of Canada (SPARCC); Wellcome Trust Case Control Consortium 2 (WTCCC2) (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat. Genet.* 45, 730–738.
- Cotsapas, C., and Hafler, D.A. (2013). Immune-mediated disease genetics: the shared basis of pathogenesis. *Trends Immunol.* 34, 22–26.

- Damoiseaux, J.G., Cautain, B., Bernard, I., Mas, M., van Breda Vriesman, P.J., Druet, P., Fournié, G., and Saoudi, A. (1999). A dominant role for the thymus and MHC genes in determining the peripheral CD4/CD8 T cell ratio in the rat. *J. Immunol.* **163**, 2983–2989.
- Deaglio, S., Dwyer, K.M., Gao, W., Friedman, D., Usheva, A., Erat, A., Chen, J.F., Enjoji, K., Linden, J., Oukka, M., et al. (2007). Adenosine generation catalyzed by CD39 and CD73 expressed on regulatory T cells mediates immune suppression. *J. Exp. Med.* **204**, 1257–1265.
- Duggal, P., Thio, C.L., Wojcik, G.L., Goedert, J.J., Mangia, A., Latanich, R., Kim, A.Y., Lauer, G.M., Chung, R.T., Peters, M.G., et al. (2013). Genome-wide association study of spontaneous resolution of hepatitis C virus infection: data from multiple cohorts. *Ann. Intern. Med.* **158**, 235–245.
- Evans, D.M., Frazer, I.H., and Martin, N.G. (1999). Genetic and environmental causes of variation in basal levels of blood cells. *Twin Res.* **2**, 250–257.
- Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zernakova, A., Stahl, E., Viatte, S., McAllister, K., et al.; Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate; Wellcome Trust Case Control Consortium (2012). High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* **44**, 1336–1340.
- Fellay, J., Ge, D., Shianna, K.V., Colombo, S., Ledergerber, B., Cirulli, E.T., Urban, T.J., Zhang, K., Gumbs, C.E., Smith, J.P., et al.; NIAID Center for HIV/AIDS Vaccine Immunology (CHAVI) (2009). Common genetic variation and the control of HIV-1 in humans. *PLoS Genet.* **5**, e1000791.
- Ferreira, M.A., Mangino, M., Brumme, C.J., Zhao, Z.Z., Medland, S.E., Wright, M.J., Nyholt, D.R., Gordon, S., Campbell, M., McEvoy, B.P., et al.; International HIV Controllers Study (2010). Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. *Am. J. Hum. Genet.* **86**, 88–92.
- Forthal, D.N., Landucci, G., Bream, J., Jacobson, L.P., Phan, T.B., and Montoya, B. (2007). FcγRIIIa genotype predicts progression of HIV infection. *J. Immunol.* **179**, 7916–7923.
- Gattinoni, L., Lugli, E., Ji, Y., Pos, Z., Paulos, C.M., Quigley, M.F., Almeida, J.R., Gostick, E., Yu, Z., Carpenito, C., et al. (2011). A human memory T cell subset with stem cell-like properties. *Nat. Med.* **17**, 1290–1297.
- Hall, M.A., Ahmadi, K.R., Norman, P., Snieder, H., MacGregor, A.J., Vaughan, R.W., Spector, T.D., and Lanchbury, J.S. (2000). Genetic influence on peripheral blood T lymphocyte levels. *Genes Immun.* **1**, 423–427.
- Harley, J.B., Alarcón-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L., Tsao, B.P., Vyse, T.J., Langefeld, C.D., Nath, S.K., et al.; International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEn) (2008). Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat. Genet.* **40**, 204–210.
- Hinks, A., Cobb, J., Marion, M.C., Prahalad, S., Sudman, M., Bowes, J., Martin, P., Comeau, M.E., Sajuthi, S., Andrews, R., et al.; Boston Children's JIA Registry; British Society of Paediatric and Adolescent Rheumatology (BSPAR) Study Group; Childhood Arthritis Prospective Study (CAPS); Childhood Arthritis Response to Medication Study (CHARMS); German Society for Pediatric Rheumatology (GKJR); JIA Gene Expression Study; NIAMS JIA Genetic Registry; TREAT Study; United Kingdom Juvenile Idiopathic Arthritis Genetics Consortium (UKJAGC) (2013). Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat. Genet.* **45**, 664–669.
- Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529.
- Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al.; International IBD Genetics Consortium (IBDGC) (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124.
- Khor, C.C., Davila, S., Breunis, W.B., Lee, Y.C., Shimizu, C., Wright, V.J., Yeung, R.S., Tan, D.E., Sim, K.S., Wang, J.J., et al.; Hong Kong–Shanghai Kawasaki Disease Genetics Consortium; Korean Kawasaki Disease Genetics Consortium; Taiwan Kawasaki Disease Genetics Consortium; International Kawasaki Disease Genetics Consortium; US Kawasaki Disease Genetics Consortium; Blue Mountains Eye Study (2011). Genome-wide association study identifies FCGR2A as a susceptibility locus for Kawasaki disease. *Nat. Genet.* **43**, 1241–1246.
- Kim, Y.J., Kim, H.Y., Lee, J.H., Yu, S.J., Yoon, J.H., Lee, H.S., Kim, C.Y., Cheong, J.Y., Cho, S.W., Park, N.H., et al. (2013). A genome-wide association study identified new variants associated with the risk of chronic hepatitis B. *Hum. Mol. Genet.* **22**, 4233–4238.
- Kirino, Y., Bertsias, G., Ishigatsubo, Y., Mizuki, N., Tugal-Tutkun, I., Seyahi, E., Ozyazgan, Y., Sacli, F.S., Erer, B., Inoko, H., et al. (2013). Genome-wide association analysis identifies new susceptibility loci for Behçet's disease and epistasis between HLA-B\*51 and ERAP1. *Nat. Genet.* **45**, 202–207.
- Kraal, G., Weissman, I.L., and Butcher, E.C. (1983). Genetic control of T-cell subset representation in inbred mice. *Immunogenetics* **18**, 585–592.
- Kyvic, K. (2000). Generalisability and assumptions of twin studies. In *Advances in Twin and Sib-Pair Analysis*, T.D. Spector, H. Sneider, and A.J. MacGregor, eds. (London: Greenwich Medical Media), pp. 67–77.
- Lugli, E., Dominguez, M.H., Gattinoni, L., Chattopadhyay, P.K., Bolton, D.L., Song, K., Klatt, N.R., Brenchley, J.M., Vaccari, M., Gostick, E., et al. (2013). Superior T memory stem cell persistence supports long-lived T cell memory. *J. Clin. Invest.* **123**, 594–599.
- Mahnke, Y.D., Beddall, M.H., and Roederer, M. (2012). OMIP-013: differentiation of human T-cells. *Cytometry A* **81**, 935–936.
- Mahnke, Y.D., Beddall, M.H., and Roederer, M. (2013a). OMIP-015: human regulatory and activated T-cells without intracellular staining. *Cytometry A* **83**, 179–181.
- Mahnke, Y.D., Beddall, M.H., and Roederer, M. (2013b). OMIP-017: human CD4(+) helper T-cell subsets including follicular helper cells. *Cytometry A* **83**, 439–440.
- Mahnke, Y.D., Beddall, M.H., and Roederer, M. (2013c). OMIP-019: quantification of human γδT-cells, iNKT-cells, and hematopoietic precursors. *Cytometry A* **83**, 676–678.
- Mbarek, H., Ochi, H., Urabe, Y., Kumar, V., Kubo, M., Hosono, N., Takahashi, A., Kamatani, Y., Miki, D., Abe, H., et al. (2011). A genome-wide association study of chronic hepatitis B identified novel risk locus in a Japanese population. *Hum. Mol. Genet.* **20**, 3884–3892.
- Moayyeri, A., Hammond, C.J., Hart, D.J., and Spector, T.D. (2012). Effects of age on genetic influence on bone loss over 17 years in women: the Healthy Ageing Twin Study (HATS). *J. Bone Miner. Res.* **27**, 2170–2178.
- Nalls, M.A., Couper, D.J., Tanaka, T., van Rooij, F.J., Chen, M.H., Smith, A.V., Toniolo, D., Zakai, N.A., Yang, Q., Greinacher, A., et al. (2011). Multiple loci are associated with white blood cell phenotypes. *PLoS Genet.* **7**, e1002113.
- Neale, M., and Cardon, L. (1992). *Methodology for Genetic Studies of Twins and Families* (Dordrecht: Kluwer Academic Publishers).
- Okada, Y., Hirota, T., Kamatani, Y., Takahashi, A., Ohmiya, H., Kumasaka, N., Higasa, K., Yamaguchi-Kabata, Y., Hosono, N., Nalls, M.A., et al. (2011). Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.* **7**, e1002067.
- Orrù, V., Steri, M., Sole, G., Sidore, C., Virdis, F., Dei, M., Lai, S., Zoledziwska, M., Busonero, F., Mulas, A., et al. (2013). Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256.
- Parkes, M., Cortes, A., van Heel, D.A., and Brown, M.A. (2013). Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **14**, 661–673.
- Pelak, K., Goldstein, D.B., Walley, N.M., Fellay, J., Ge, D., Shianna, K.V., Gumbs, C., Gao, X., Maia, J.M., Cronin, K.D., et al.; Infectious Disease Clinical Research Program HIV Working Group; National Institute of Allergy and Infectious Diseases Center for HIV/AIDS Vaccine Immunology (CHAVI) (2010). Host determinants of HIV-1 control in African Americans. *J. Infect. Dis.* **201**, 1141–1149.
- Pereyra, F., Jia, X., McLaren, P.J., Telenti, A., de Bakker, P.I., Walker, B.D., Ripke, S., Brumme, C.J., Pulit, S.L., Carrington, M., et al. (2010). The major



genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330, 1551–1557.

Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinowski, N., Frohlich, I., et al. (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* 344, 519–523.

Raulet, D.H., Gasser, S., Gowen, B.G., Deng, W., and Jung, H. (2013). Regulation of ligands for the NKG2D activating receptor. *Annu. Rev. Immunol.* 31, 413–441.

Richards, J.B., Rivadeneira, F., Inouye, M., Pastinen, T.M., Soranzo, N., Wilson, S.G., Andrew, T., Falchi, M., Gwilliam, R., Ahmadi, K.R., et al. (2008). Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* 371, 1505–1512.

Roederer, M., Nozzi, J.L., and Nason, M.C. (2011). SPICE: exploration and analysis of post-cytometric complex multivariate datasets. *Cytometry A* 79, 167–174.

Sakaguchi, S., Miyara, M., Costantino, C.M., and Hafler, D.A. (2010). FOXP3+ regulatory T cells in the human immune system. *Nat. Rev. Immunol.* 10, 490–500.

Sinha, S., Mishra, S.K., Sharma, S., Patibandla, P.K., Mallick, P.K., Sharma, S.K., Mohanty, S., Pati, S.S., Mishra, S.K., Ramteke, B.K., et al.; Indian Genome Variation Consortium (2008). Polymorphisms of TNF-enhancer and gene for FcγRIIIa correlate with the severity of falciparum malaria in the ethnically diverse Indian population. *Malar. J.* 7, 13.

Soranzo, N., Rivadeneira, F., Chinappan-Horsley, U., Malkina, I., Richards, J.B., Hammond, N., Stolk, L., Nica, A., Inouye, M., Hofman, A., et al. (2009). Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS Genet.* 5, e1000445.

Starbeck-Miller, G.R., Badovinac, V.P., Barber, D.L., and Harty, J.T. (2014). Cutting edge: Expression of FcγRIIB tempers memory CD8 T cell function in vivo. *J. Immunol.* 192, 35–39.

Strasser, A., Jost, P.J., and Nagata, S. (2009). The many roles of FAS receptor signaling in the immune system. *Immunity* 30, 180–192.

Teo, Y.Y., Inouye, M., Small, K.S., Gwilliam, R., Deloukas, P., Kwiatkowski, D.P., and Clark, T.G. (2007). A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 23, 2741–2746.

Thye, T., Vannberg, F.O., Wong, S.H., Owusu-Dabo, E., Osei, I., Gyapong, J., Sirugo, G., Sisay-Joof, F., Enimil, A., Chinbuah, M.A., et al.; African TB Genetics Consortium; Wellcome Trust Case Control Consortium (2010). Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* 42, 739–741.

Urabe, Y., Ochi, H., Kato, N., Kumar, V., Takahashi, A., Muroyama, R., Hosono, N., Otsuka, M., Tateishi, R., Lo, P.H., et al. (2013). A genome-wide association study of HCV-induced liver cirrhosis in the Japanese population identifies novel susceptibility loci at the MHC region. *J. Hepatol.* 58, 875–882.

van Dongen, J., Slagboom, P.E., Draisma, H.H., Martin, N.G., and Boomsma, D.I. (2012). The continuing value of twin studies in the omics era. *Nat. Rev. Genet.* 13, 640–653.

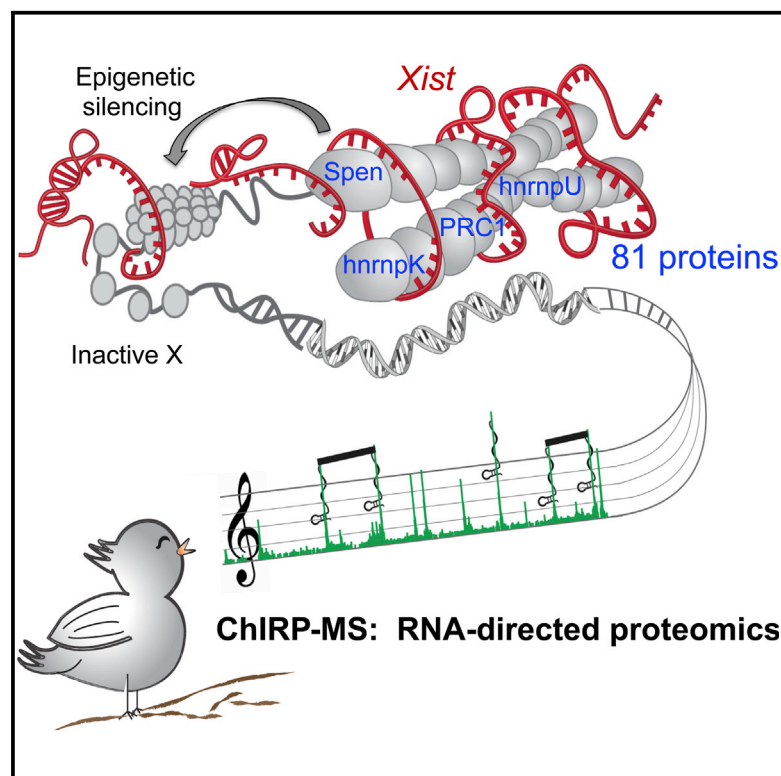
Wang, S.S., Cerhan, J.R., Hartge, P., Davis, S., Cozen, W., Severson, R.K., Chatterjee, N., Yeager, M., Chanock, S.J., and Rothman, N. (2006). Common genetic variants in proinflammatory and other immunoregulatory genes and risk for non-Hodgkin lymphoma. *Cancer Res.* 66, 9771–9780.

Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26, 2190–2191.

Zhang, F., Liu, H., Chen, S., Low, H., Sun, L., Cui, Y., Chu, T., Li, Y., Fu, X., Yu, Y., et al. (2011). Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat. Genet.* 43, 1247–1251.

# Systematic Discovery of Xist RNA Binding Proteins

## Graphical Abstract



## Authors

Ci Chu, Qiangfeng Cliff Zhang, ..., Edith Heard, Howard Y. Chang

## Correspondence

howchang@stanford.edu

## In Brief

Development of a general method for identifying RNA-protein interactions in vivo reveals 81 endogenous proteins that associate with Xist RNA in two waves to control mammalian dosage compensation.

## Highlights

- ChIRP-MS identifies endogenous protein partners associated with specific RNAs
- Xist lncRNA binds 81 proteins in two waves during X chromosome inactivation
- HnRnpK is required for Xist-mediated chromatin modifications and Polycomb targeting
- Spen binds to Xist A-repeat and is required for gene silencing



# Systematic Discovery of Xist RNA Binding Proteins

Ci Chu,<sup>1,2</sup> Qiangfeng Cliff Zhang,<sup>1</sup> Simão Teixeira da Rocha,<sup>3</sup> Ryan A. Flynn,<sup>1</sup> Maheetha Bharadwaj,<sup>1</sup> J. Mauro Calabrese,<sup>4</sup> Terry Magnuson,<sup>5</sup> Edith Heard,<sup>3</sup> and Howard Y. Chang<sup>1,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute and Program in Epithelial Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

<sup>2</sup>Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Singapore

<sup>3</sup>Institut Curie, CNRS UMR3215, INSERM U934, 26 rue d'Ulm, Paris 75248, France

<sup>4</sup>Department of Pharmacology and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA

<sup>5</sup>Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA

\*Correspondence: [howchang@stanford.edu](mailto:howchang@stanford.edu)

<http://dx.doi.org/10.1016/j.cell.2015.03.025>

## SUMMARY

Noncoding RNAs (ncRNAs) function with associated proteins to effect complex structural and regulatory outcomes. To reveal the composition and dynamics of specific noncoding RNA-protein complexes (RNPs) in vivo, we developed comprehensive identification of RNA binding proteins by mass spectrometry (ChIRP-MS). ChIRP-MS analysis of four ncRNAs captures key protein interactors, including a U1-specific link to the 3' RNA processing machinery. Xist, an essential lncRNA for X chromosome inactivation (XCI), interacts with 81 proteins from chromatin modification, nuclear matrix, and RNA remodeling pathways. The Xist RNA-protein particle assembles in two steps coupled with the transition from pluripotency to differentiation. Specific interactors include HnmpK, which participates in Xist-mediated gene silencing and histone modifications but not Xist localization, and *Drosophila* Split ends homolog Spen, which interacts via the A-repeat domain of Xist and is required for gene silencing. Thus, Xist lncRNA engages with proteins in a modular and developmentally controlled manner to coordinate chromatin spreading and silencing.

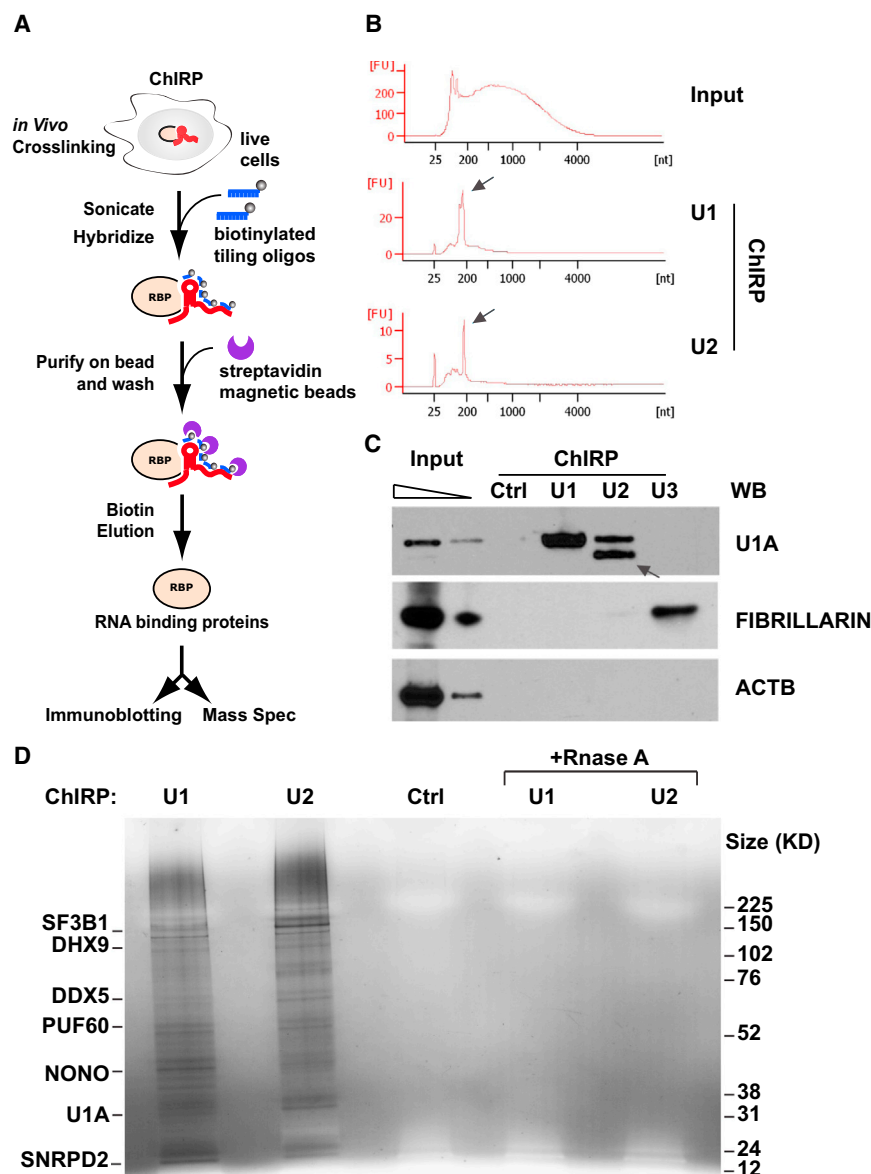
## INTRODUCTION

Many long noncoding RNAs (lncRNAs) are recently recognized as functional regulators of gene expression (Rinn and Chang, 2012), but their mechanisms of action are largely unknown. RNA binding proteins (RBPs) play key roles in lncRNA-mediated gene regulation, and obtaining the full interaction map of proteins bound to a lncRNA of interest is critical to our understanding of its function. Many tools have been developed to describe RNA-protein interaction from a protein-centric view, typically by immunoprecipitating a protein and analyzing the associated RNAs with a microarray or high-throughput sequencing (reviewed by Riley and Steitz, 2013). In contrast, fewer methods are available from the perspective of a particular RNA. This is usually achieved by (1) tagging the RNA with affinity aptamers,

which involves complicated genetic engineering; (2) using in-vitro-transcribed RNA to retrieve proteins from native cell lysates (RNA chromatography), which is prone to the formation of non-physiological RNA-protein interactions; and (3) using immobilized oligonucleotides to capture RNA:protein complex under native conditions, which suffers from both post-lysis re-associations and unpredictable specificity of target RNA retrieval (reviewed by Chu et al., 2015). The ideal strategy should capture in vivo lncRNA-protein interactions, achieve high yield and specificity without genetic tagging, and provide comprehensive portraits of lncRNP in diverse biological states.

Xist is a lncRNA (17 kb long in the mouse) required for X chromosome inactivation (XCI) of one of the two X chromosomes in female cells, thus enabling dosage compensation between XX females and XY males (Gendrel and Heard, 2011). XCI takes place early in embryonic development and is thought to occur in multiple steps: counting and choosing the X chromosome to silence, spreading of Xist over the target X chromosome, and silencing of most of its active genes (Payer and Lee, 2008). The latter two steps are believed to be mediated by specific Xist-associated protein factors, which remain largely mysterious. Xist expression marks the future inactive X chromosome (Xi) and is sufficient to recruit silencing chromatin modification complex such as the Polycomb proteins (Gendrel and Heard, 2011). It has been debated whether Xist RNA physically recruits one or more silencing factors or whether Xist indirectly promotes transcriptional silencing via reinforcement of repressive chromatin. XCI is also developmentally regulated in several important ways. In the mouse, XCI can proceed by random inactivation of either paternal or maternal chromosome in somatic cells or by always inactivating the paternally derived X in extra-embryonic cells, a process called imprinted XCI (Takagi and Sasaki, 1975). During random XCI, Xist is not expressed in pluripotent embryonic stem cells (ESCs) and is upregulated during differentiation (Wutz and Jaenisch, 2000). Ectopic Xist RNA coating can induce gene silencing in ESCs, although this is reversible during an early differentiation time window, becoming irreversible at later stages (Wutz and Jaenisch, 2000). Knowledge of the Xist lncRNP in these diverse states may provide insights into this classic and intricate epigenetic system.

Here, we introduce comprehensive identification of RBPs by mass spectrometry (ChIRP-MS), an optimized method for the identification of lncRNA-bound proteome. Applying ChIRP-MS to four noncoding RNAs, we found known and validated novel



**Figure 1. ChIRP-MS Method and Validation**

(A) Outline of the ChIRP-MS workflow. Briefly, RNP complexes are crosslinked *in vivo* by 3% formaldehyde for 30 min and solubilized by sonication. Target ncRNA are pulled out by biotinylated antisense oligos, and associated proteins are eluted with free biotin, separated by electrophoresis. Each size fraction is subjected to LC/MS-MS identification.

(B) Distribution of input and U1- and U2-enriched RNA sizes, as determined by Bioanalyzer (Agilent). (C) Proteins retrieved by U1, U2, U3, and control probes, analyzed by immunoblotting. Arrow indicates the U1A close homolog, U2B, cross-identified by U1A antibody.

(D) Proteins retrieved by U1, U2, non-targeting probe control, and RNase-treated controls, visualized by silver staining. Major proteins enriched are indicated on the left.

matography-tandem mass spectrometry (LC-MS/MS). We conducted negative controls by use of non-interacting control probes, RNase treatment of lysate prior to ChIRP, or genetic removal of the target RNA.

As a proof of principle, we performed ChIRP-MS of human U1 and U2 snRNAs in HeLa S3 cells. The snRNAs are ideal for validating ChIRP-MS because they are abundant (~1 million copies of U1 per cell) (Gesteland and Atkins, 1993) and the spliceosome composition is well known (Pomeranz Krummel et al., 2009; Stark et al., 2001; Zhou et al., 2002). Furthermore, non-canonical roles of U1 in preventing premature mRNA cleavage and polyadenylation have been recently reported (Almada et al., 2013; Berg et al., 2012; Kaida et al., 2010), implying potential novel interactors that ChIRP-MS may discover. We designed anti-

functional interactors. By performing Xist ChIRP-MS in different cell states, lineages, and cell types and with mutant Xist alleles, we uncover mechanisms of dynamic and coordinate assembly of Xist binding partners, suggesting an organizing principle for lncRNPs.

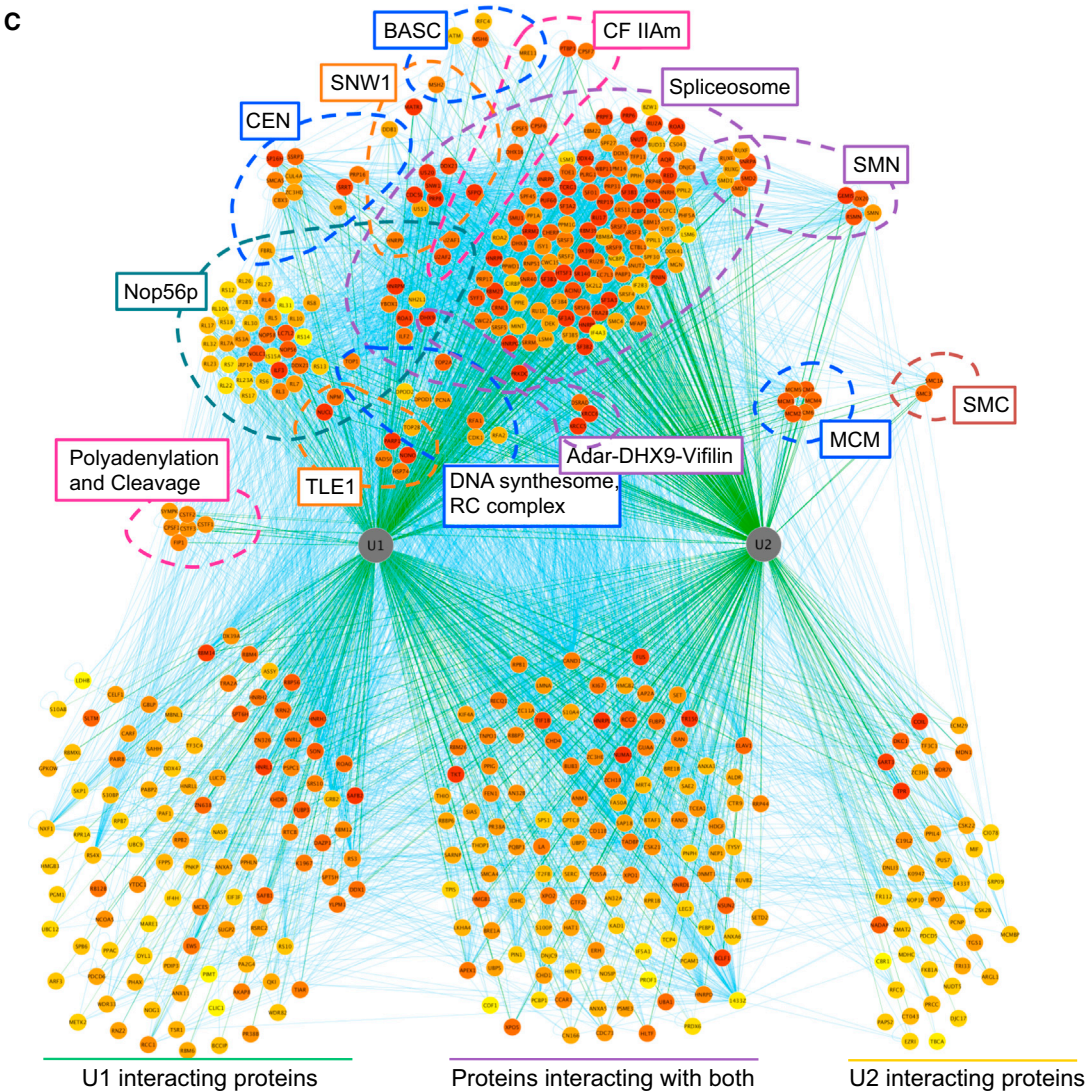
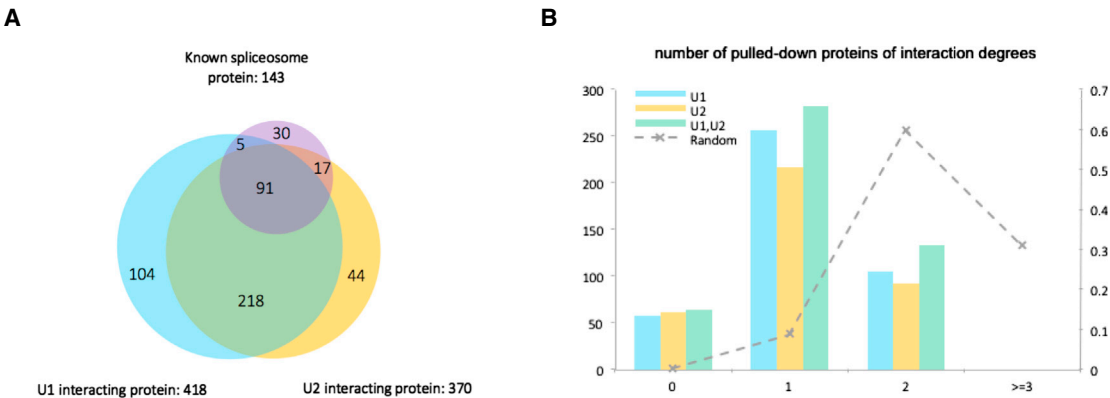
## RESULTS

### ChIRP-MS Method

Extending on ChIRP-seq, a method using DNA oligonucleotides to capture lncRNAs and their genomic DNA binding sites (Chu et al., 2011), we optimized ChIRP-MS to identify lncRNA-associated proteins (Figure 1A). We cross-link cells extensively with formaldehyde, retrieve target RNA with oligonucleotide hybridization, and use a gentle biotin-elution to liberate associated proteins. The enriched proteins were identified by liquid chro-

matography-tandem mass spectrometry (LC-MS/MS). We conducted negative controls by use of non-interacting control probes, RNase treatment of lysate prior to ChIRP, or genetic removal of the target RNA. As a proof of principle, we performed ChIRP-MS of human U1 and U2 snRNAs in HeLa S3 cells. The snRNAs are ideal for validating ChIRP-MS because they are abundant (~1 million copies of U1 per cell) (Gesteland and Atkins, 1993) and the spliceosome composition is well known (Pomeranz Krummel et al., 2009; Stark et al., 2001; Zhou et al., 2002). Furthermore, non-canonical roles of U1 in preventing premature mRNA cleavage and polyadenylation have been recently reported (Almada et al., 2013; Berg et al., 2012; Kaida et al., 2010), implying potential novel interactors that ChIRP-MS may discover. We designed anti-





(legend on next page)

serving as another negative control. These results indicate that ChIRP is specific even for very abundant RBPs.

### U1 and U2 ChIRP-MS Reveal Known and Novel Interactors

We next scaled up experiments for MS-level analysis, including both RNase and non-targeting probe controls. Silver staining of ChIRP samples showed that U1 and U2 probes pulled down rich proteins from HeLa lysates, whereas all control samples are clean (Figure 1D), indicating that ChIRP-MS is highly specific on the proteome level. U1 and U2 ChIRP-MS enriched (by  $>\log_2 3.5$  or  $>10$ -fold, see Experimental Procedures) more than 400 proteins over respective negative controls (Figure 2A, full peptide count list in Table S1). The results were highly reproducible regardless of control strategies: for U1, 98% overlap between RNase and non-targeting probe controls; 99% for U2. The near-identical results from using two orthogonal methods for background removal highlight the robustness of the protocol.

U1 and U2 snRNAs shared their RBPs extensively (309 in common, or 74% of U1 and 84% of U2-RBPs), as predicted from their common cellular function (Figure 2A). Both U1 and U2 strongly enriched for proteins involved in splicing and pre-mRNA biogenesis, as anticipated (Figure S1A). Together, the two snRNAs retrieved 79% of the human spliceosome components (Figure 2A) and 8 of 9 direct U1 binding proteins verified by crystal structure (Pomeranz Krummel et al., 2009; Stark et al., 2001; Ruepp et al., 2008). Analysis of known protein-protein interaction networks showed that the vast majority (96%) of all proteins identified were within two degrees of separations from the core spliceosome (Figure 2B) or the direct binding proteins of U1 (Figure S2A) (Pomeranz Krummel et al., 2009; Zhang et al., 2012; Stark et al., 2001), suggesting that ChIRP-MS yields the immediate and most relevant protein network. Organization of U1/U2 interactomes into complexes based on curated protein interaction data confirmed extensive coverage of the spliceosome, SMN, and cap binding complexes (Figure 2C).

U1 selectively enriched for the CSTF complex involved in pre-mRNA cleavage and polyadenylation, a recently described non-canonical function of U1 (Figures 2C and S2B; Gene Ontology (GO) term “RNA 3'-end processing” in Figure S1A) (Berg et al., 2012; Kaida et al., 2010). Immunoblots validated U1-selective pull-down of CSTF2 over other snRNAs, which potentially explains this U1-exclusive function (Figure S1B) (Berg et al., 2012) and shows that ChIRP is specific for proximal interactions even within the same complex (e.g., the spliceosome). These and other protein complexes discovered represent a wealth of information for the snRNP community (Figures 2C and S1A).

### Xist Ribonucleoprotein Complex Purification

We next turned to discover the protein partners of Xist. ChIRP-MS of Xist represents a substantial challenge in several ways: (1) Xist is far less abundant than U1 ( $<2,000$  copies per cell versus 1 million) (Buzin et al., 1994), making it more relevant to other regulatory lncRNAs; (2) Xist transcript is long and will be sheared into fragments, requiring a tiling-probe strategy not necessary for the study of U1/U2; (3) Xist is chromatin and nuclear matrix associated and therefore insoluble even by detergent and nuclease extraction (Clemson et al., 1996). Based on these considerations, we designed 43 probes against the mouse Xist RNA (Table S2). In a female mouse cell line (Neuro2a), we confirmed that Xist RNA was completely solubilized by sonication (data not shown), and over 60% of Xist RNA was selectively retrieved without enrichment of housekeeping *Gapdh* mRNA (Figure 3A).

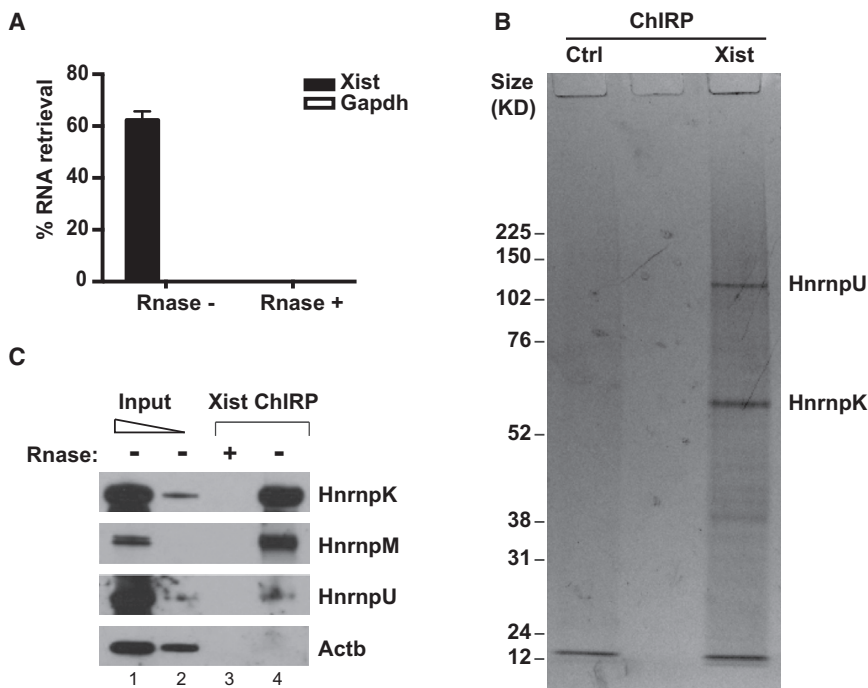
Xist probes retrieved rich protein analytes compared to the RNase control (Figure 3B). The most abundant proteins retrieved are HnrnpK and U, and M, the first two readily visualizable by Coomassie blue (Figure 3B). HnrnpU is required for the spread of Xist RNA across the chromosome in *cis* (Hasegawa et al., 2010), thus a positive control. Xist-dependent retrieval of all three proteins was validated by ChIRP-western, proving that they are not retrieved by virtue of their sheer abundance; the control protein beta-actin was not enriched (Figure 3C).

### Stepwise and Developmentally Regulated Assembly of Xist RNP

We carefully selected biological systems to perform Xist ChIRP-MS that represents different stages of Xist-mediated silencing (Figure 4A). Although Xist is expressed in most differentiated female cells, it is largely dispensable for the maintenance of XCI (Brown and Willard, 1994; Csankovszki et al., 1999). To ensure that we catch Xist “in action,” we chose a male mouse ESC line that has been genetically engineered to harbor a Xist cDNA knocked into chromosome 11 (chr11) that is inducible by doxycycline (dox) (Wutz and Jaenisch, 2000). The exogenous Xist localizes to chr11 and silences chr11 genes at a long distance after 4 days of sustained expression and retinoic acid-induced differentiation (Wutz and Jaenisch, 2000) (Figure 4A, lanes 1 and 2). Turning on or off Xist transcription with dox creates an isogenically controlled experiment. Furthermore, the relatively rapid initiation of Xist silencing ensures synchronicity among cells, suppressing noise arising from population heterogeneity. To study the endogenous Xist lncRNP, we performed parallel ChIRP-MS in an epiblast stem cell line (EpiSC) (Gillich et al., 2012). EpiSCs are derived from E5.5–E6.5 epiblasts and represent cells that have just undergone random XCI (occurring  $\sim$ E5.5) (Hayashi and Surani, 2009; Rastan, 1982; Takagi et al., 1982) (Figure 4A, lane 3). Finally, we performed Xist ChIRP-MS

#### Figure 2. U1/U2 ChIRP-MS

- (A) Venn diagram of known spliceosome proteins and proteins pulled down by U1 or U2. The number of interactions in each set is given after the set label.  
 (B) Numbers of U1/U2 pulled-down proteins by their degrees of separation from known spliceosome proteins. The dashed line represents the distribution of a randomly simulated set of the same number of proteins pulled down by U1 and U2 (right axis).  
 (C) Protein-protein and protein-RNA interaction network of U1/U2 pulled-down proteins. Proteins belonging to known complexes are organized and annotated in groups in top half of the plot, and proteins of unknown affiliation are presented at the bottom. Complexes and proteins more strongly enriched by U1 (left in graph) (e.g., Polyadenylation and cleavage, Nop56p) are positioned accordingly.



**Figure 3. Xist ChIRP-MS**

(A) >60% of Xist RNA was retrieved from the cell by ChIRP, while no Gapdh was detected. RNase treatment eliminates Xist transcripts prior to pull-down.

(B) Proteins retrieved by Xist and isogenic control (no Xist) visualized by Coomassie blue staining.

(C) Validation of ChIRP-enriched proteins by immunoblotting.

lapping proteins (8/14) are “nonspecific ChIRP hits” as defined above.

Xist ChIRP-MS in all four cell types retrieved a common set of proteins (62/81, 77%), termed Set 1. An additional 19 proteins interacted with Xist only in differentiated ESC, EpiSC, and TSCs; these proteins are termed Set 2. We describe the identity of proteins in these two sets and then discuss the dynamics of the interactions. Some of the binding proteins were known factors involved in XCI. We identified Rnf2 (also known as Ring2 or Ring1b), the catalytic subunit of Polycomb repressive complex 1 (PRC1) that

in trophoblast stem cells (TSCs), where the paternal X chromosome is always silenced (Calabrese et al., 2012), a phenomenon termed imprinted XCI that contrasts with the random XCI in somatic cells (Figure 4A, lane 4). RNase controls were performed side by side in the EpiSC and TSC experiments.

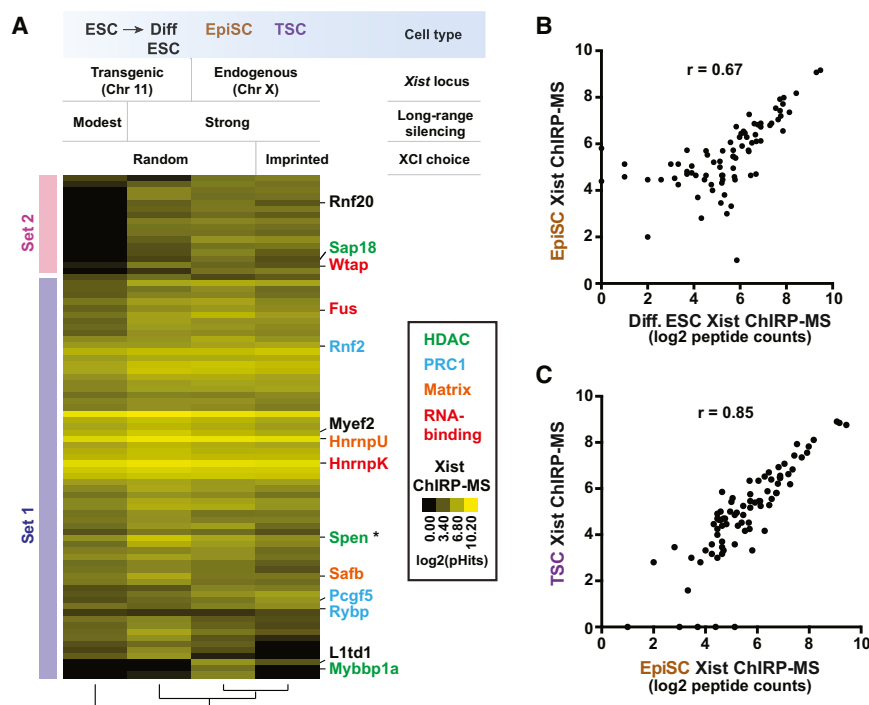
We also compared Xist ChIRP-MS to ChIRP-MS of three abundant nuclear RNAs—U1, U2, and 7SK—to evaluate Xist-specific interactions (Experimental Procedures). 7SK is a snRNA present at ~200,000 copies per cell and is involved in transcriptional elongation control. We ranked peptides enriched by each ncRNA and prioritized proteins that had Xist ChIRP-MS enrichment ranking at least 2-fold better than rankings in any of the three comparator ncRNAs.

In total, we identified 81 Xist binding proteins from the four experiments (Figure 4A and full list of enriched proteins with peptide counts reported in Table S3). When compared to U1/U2/7SK, only a minority of Xist hits (30/81) was also highly enriched by another ncRNA (rank ratio < 2, Table S4). These non-specific proteins are mainly involved in RNA processing (GO enrichment  $p = 8.4E-28$ ) and may be involved in nuclear ncRNA splicing, nuclear retention, or stability. They are likely bona-fide Xist binding proteins because they pass RNase and genetic controls, but they may not contribute to the specific gene regulatory function of Xist. We provide the list of non-specific proteins retrieved by all four nuclear ncRNAs as a resource for the field (in red, Table S4). In contrast, the Xist-specific proteins selectively enriched for gene repressors (GO enrichment  $p = 9.6E-8$ ), which are highlighted in Table S4 and discussed below. We also overlapped the set of proteins retrieved by Xist with those retrieved by two other abundant nuclear lncRNAs, NEAT1 and MALAT1, and found limited overlap (14 out of 81 shared by all three, Figure S2C) (West et al., 2014). As expected, the majority of over-

deposits the repressive lysine119 monoubiquitination on histone H2A (H2AK119ub) over the inactive X chromosome (de Napoles et al., 2004; Fang et al., 2004). Other PRC1 components identified included Pcgf5 and Rybp (both in set 1); Rybp is a stoichiometric component of PRC1 that has been shown to accumulate on the Xi independently of PRC2 (Tavares et al., 2012). We also found the Sin3-HDAC1 components Spen, Sap18, and Mybbp1a, which are repressive transcriptional factors that recruit histone deacetylase (HDAC) complexes. Histone deacetylation correlates with reduced gene expression and is another hallmark of the inactive X chromosome (Keohane et al., 1996). The co-purification of these proteins may bridge the biochemical gap between Xist and HDAC that remains little explored in the field. Xist ChIRP-MS also recovered nuclear matrix proteins HnrnpU, Matrin 3, and Safb, consistent with the observation that Xist is probably anchored by nuclear matrix (Clemson et al., 1996). Notably, HnrnpU is required for Xist localization (Hasegawa et al., 2010). Finally, RBPs such as HnrnpK strongly and specifically interacted with Xist; HnrnpK was not retrieved by U1 or U2. Collectively, the two sets of proteins represent candidate factors that could play roles in Xist localization or function.

Comparison of Xist interactors in the four cell types revealed a potential step-wise assembly of Xist binding proteins from the pluripotent state to differentiation. Unsupervised hierarchical analysis showed that the Xist interactors are distinct in ESCs, whereas the differentiated ESC, EpiSC, and TSC shared a significant degree of overlap (Figure 4A). Whereas Set 1 proteins remain associated with Xist from pluripotency to differentiation, Xist interaction with Set 2 proteins is observed only upon differentiation. Xist interacted with both Set 1 and Set 2 proteins in the latter three cell types; 77 of 81 Xist interactors (95%) were independently retrieved in these differentiated cells. The HDAC





**Figure 4. Xist Partner Proteins Are Developmentally Regulated**

(A) Heatmap of Xist-RBPs pulled down in the four experiments. Color bars indicate abundance of peptides detected. Protein annotations were color designated based on their class. (B and C) (B) Similar proteins are enriched between differentiating ES cells versus EpiSCs and (C) between EpiSCs and TSCs.

complex subunit Spen straddles these categories because it interacts with Xist in ESC, but the interaction intensifies with differentiation (asterisk in Figure 4A). The distinction between Set 1 and 2 is unlikely due to lower efficiency of Xist ChIRP-MS in ESCs because the quantitative signal for Set1 proteins in ESC is on par with that in differentiated cells. While the Set 1 proteins may represent the ground state of Xist-interactome that prepares the lncRNA for action, the differentiation-coupled Xist interactors include intriguing chromatin-modifying proteins such as Spen, Rnf20, Mybbp1a, and Sap18. These may represent additional silencing factors recruited to Xist RNA when XCI is in full action. Quantitative comparison between Xist ChIRP-MS in differentiated ESC versus EpiSC or versus TSC showed that they are largely similar, especially for the strong interactors ( $r = 0.67$  and  $0.85$ , respectively, Figures 4B and 4C). These results suggest that (1) transgenic Xist indeed phenocopies the endogenous RNA and shares similar binding proteins; (2) ChIRP-MS is robust and gives consistent results in multiple systems; (3) random XCI and imprinted XCI appear to employ nearly identical Xist-associated proteins, and therefore, extraembryonic trophoctoderm likely executes silencing in ways that are highly similar to that of the embryo proper.

### HnrnpK Participates in Xist-Mediated Gene Silencing

To assess the functional importance of Xist-interacting proteins in gene silencing, we tested their dispensability in Xist-mediated silencing of the imprinted *Grb10/Meg1* gene, previously shown to be silenced by Xist upon differentiation of ESC (Wutz and Jaenisch, 2000). The imprinted *Grb10* gene is located 41 megabases away from the Xist transgene on chr11 and is thought to be mono-allelically expressed from the chr11 harboring the transgene (Figure 5A) (Wutz and Jaenisch, 2000). We showed it indeed was

silenced by transgenic Xist (Figure 5B). We chose to first target HnrnpK, M, and U because they represent some of the most enriched Set 1 proteins (especially K) and because this simple heuristic identifies HnrnpU, a known key mediator of Xist function. Upon siRNA-mediated depletion (Figure 5C), only HnrnpU and HnrnpK had significant effects on *Grb10* silencing (Figure 5D). We ruled out off-target effects by showing that all four individual siRNAs against HnrnpK produced the same de-repression effect (Figure S4A). We directly visualized transcription from the Xist-silenced allele using two-color RNA-FISH

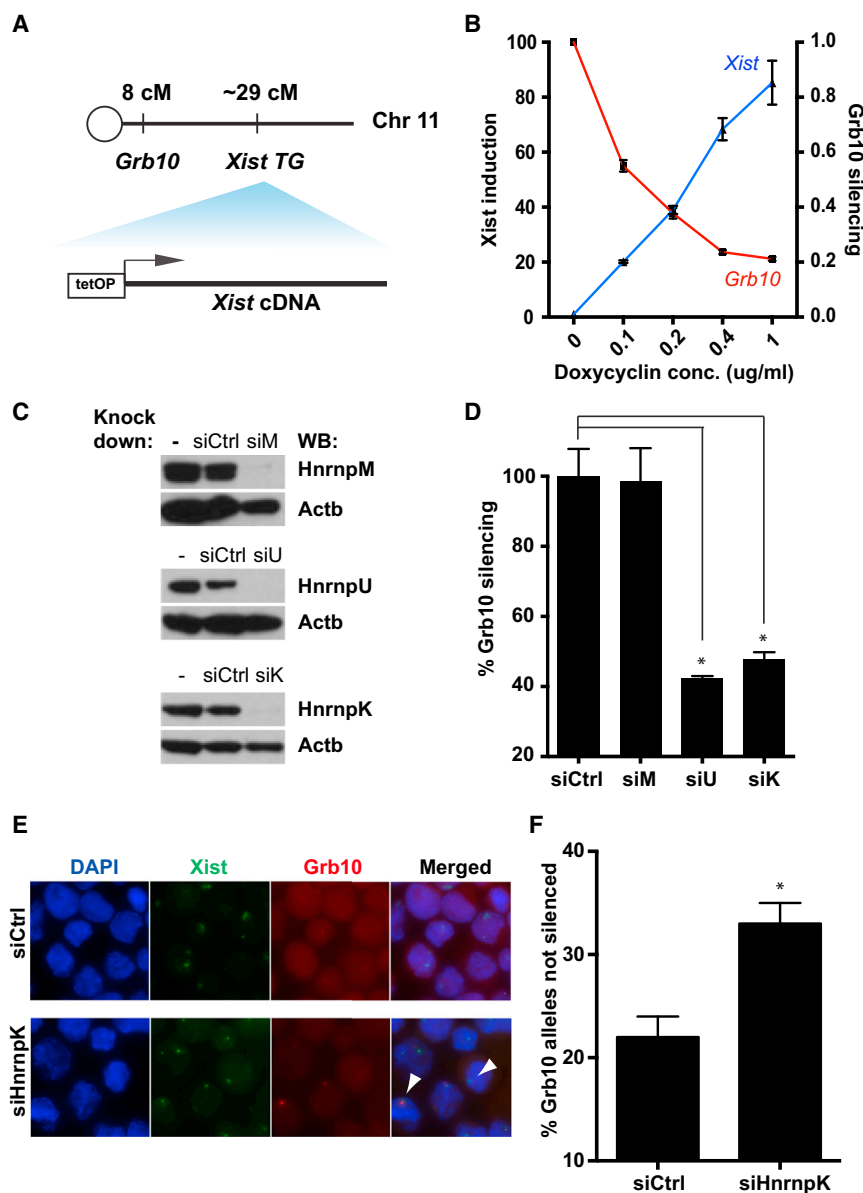
(Figures 5E and 5F). We used a genomic (BAC) probe, allowing us to detect the *Grb10* nascent transcript rather than its mature mRNA. In this way, we scored for the presence or absence of *Grb10* transcription adjacent to the Xist-coated chr11. HnrnpK depletion significantly increased the frequency of active *Grb10* allele found close to or within the Xist RNA-coated chromosome 11, indicating that, indeed, it is less sensitive to Xist-mediated silencing in the absence of HnrnpK (Figures 5E and 5F).

We also tested the requirement for HnrnpK in endogenous XCI in EpiSC. We converted ESC into EpiSC in the presence of Fgf2 and Activin (Guo et al., 2009). EpiSC conversion was confirmed by morphologic changes, marker expression, induction of Xist expression, and Xist localization to the Xi (Figures S3A–S3D). We performed single-molecule fluorescent in situ hybridization (sm-FISH) on *Usp9x*, an X-linked gene that is subject to random X-inactivation. We used FISH probes against the introns of *Usp9x* gene to exclusively detect its pre-mRNA that indicates active transcription. Although only 10% of the cells show two *Usp9x* pinpoints in control cells, HnrnpK- or HnrnpU-depleted cells showed a 2- to 3-fold increase in cells with two *Usp9x* FISH signals (Figures S3E and S3F). The reduction in successful XCI for HnrnpU depletion matched observations from a prior study (Figure S3F) (Hasegawa et al., 2010). We conclude that HnrnpK is an important factor for Xist-mediated silencing.

### HnrnpK Contributes to Xist-Mediated Chromatin Modifications, but Not Xist Biogenesis or Localization

We tested potential roles of HnrnpK early in the sequence of repressive events, including Xist biogenesis, localization, and spreading or chromatin silencing. Northern blot analysis showed that Xist abundance or splicing were not impacted by depletion of HnrnpK, U, or M (the two minor isoforms upon HnrnpU





**Figure 5. Functional Characterization of Xist RBPs**

(A) Relative positions of Grb10 and Xist transgene (TG) on chr11. (B) Induction of Xist and repression of Grb10 by different doses of dox in e36 cells that have undergone RA-induced differentiation for 4 days. (C) Western validation of HnrnpU, K, and M knockdown by siRNAs. (D) De-repression of Grb10 upon depletion of HnrnpU, K, and M. (E) Dual-color FISH of Grb10 and Xist in e36 cells that are depleted of HnrnpK. Arrowheads indicate Grb10 allele escaping Xist silencing. (F) Quantification of cells with Grb10 expression on the Xist-coated chromosome by counting >150 cells from 3 replicates.

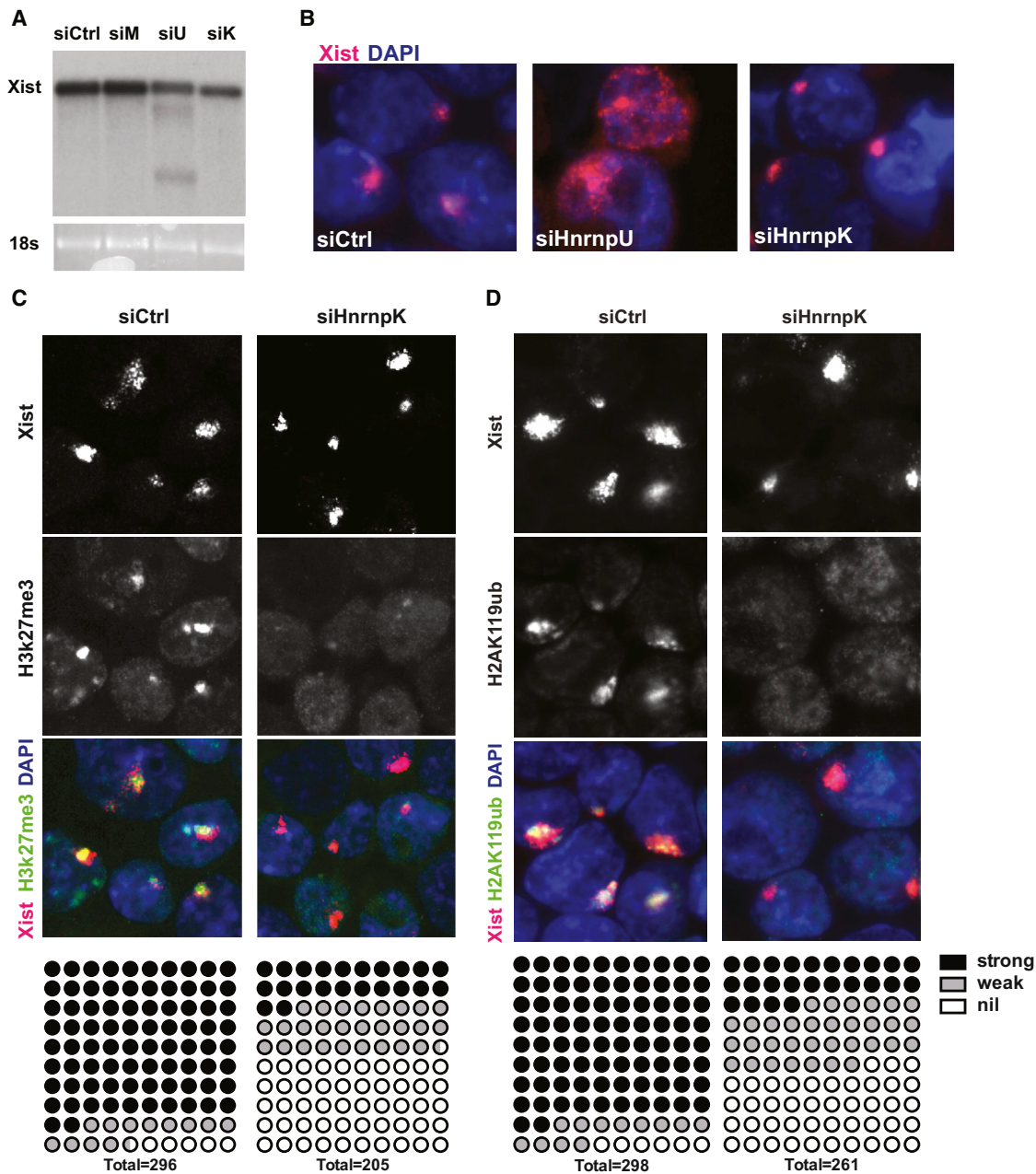
hypothesis that HnrnpK is a novel regulator of the initiation of X-inactivation. Xist ChIRP retrieved multiple PRC1 subunits, and PRC1 or PRC2 action can mutually recruit each other (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014). Indeed, HnrnpK depletion also spatially dissociated Xist from the PRC2 subunit Eed (Figure S5). HnrnpK contains three RNA-binding KH domains that may directly bind Xist. UV-cross-linking RNP immunoprecipitation followed by RT-PCR (CLIP-qRT-PCR) showed that HnrnpK directly bound Xist RNA, with the strongest interaction mapping downstream of repeat F in exon 1 (Figure S4C). HnrnpK retrieved Xist more efficiently in CLIP than HnrnpU, a known direct interaction that we reproduced (Hasegawa et al., 2010).

#### The A-Repeat of Xist Interacts with Spen to Mediate Gene Silencing

We next explored the use of ChIRP-MS to dissect domain-specific interactions of

depletion are consistent with previous report) (Hasegawa et al., 2010) (Figure 6A), although we cannot exclude that minor changes occurred given that Xist is present in multiple isoforms. Next, sm-FISH confirmed that HnrnpU depletion indeed delocalized Xist, but HnrnpK depletion did not (Figure 6B). Combined immunofluorescence and RNA FISH (IF-coFISH) showed that, while Xist RNA colocalized with H2AK119ub and H3K27me3, HnrnpK depletion significantly reduced the accumulation of H2AK119ub and H3K27me3 on the Xi without affecting Xist RNA localization (Figures 6C and 6D). HnrnpK depletion did not affect the global level of H3K27me3, showing that HnrnpK has a specific impact on Xist-mediated recruitment of repressive chromatin marks (Figure S4B). Given that both H3K27me3 and H2AK119ub modifications are among the earliest epigenetic changes occurring to the Xi, the results are consistent with our

Xist with its partner proteins. A small 0.9 kb region on the very 5' end of Xist that harbors the conserved A-repeat element is required for transcriptional silencing, but not for chromatin interaction or spreading across the X chromosome (da Rocha et al., 2014; Wutz et al., 2002). In principle, deletion of A-repeat may alter RNA folding or modification to abrogate interaction of most of the silencing proteins; alternatively, the A-repeat may be selectively required for the interaction of a small number of key silencing factors. ChIRP-MS appears to be an ideal approach to distinguish between these models. Xist ChIRP-MS of ES cells harboring inducible full-length Xist or A-repeat mutant at the endogenous locus on the X chromosome (Wutz et al., 2002) revealed that most protein interactions were not affected by the deletion, but three proteins—Spen, Rnf20, and Wtap—were completely unable to bind the mutant (Figure 7A). Notably,



**Figure 6. HnrnpK Is Required for Repressive Chromatin Modifications of Inactive X**

(A) Northern blot against Xist in e36 cells depleted of HnrnpM, U, or K.

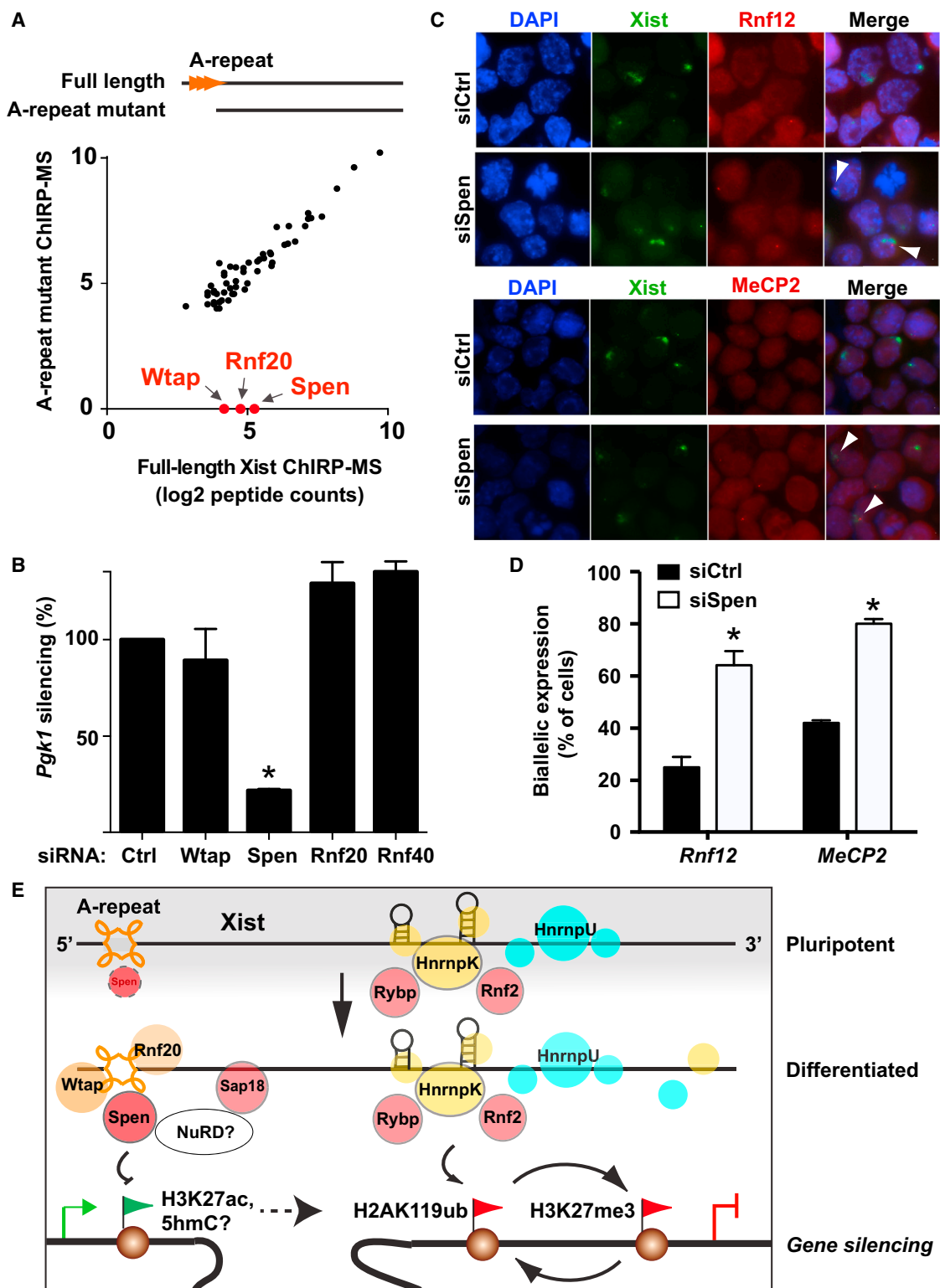
(B) Xist sm-FISH in HnrnpU and K knockdown cells.

(C and D) IF co-FISH of Xist and H3K27me3 (C)/H2AK119ub (D) in HnrnpK knockdown cells. Number of cells with strong, weak and undetectable repressive marks overlapping with Xist foci were tallied and represented below.

Spn interaction with Xist is increased upon ESC differentiation, and Rnf20 and Wtap both belong to Set 2 proteins that interact with Xist only upon differentiation (Figure 4A). Thus, the A-repeat appears to be a focus of the differentiation-coupled assembly of Xist RNP. The exclusive binding of these three proteins to full-length Xist, but not the A-repeat mutant, was confirmed by ChIRP-western (Figure S6A). This result also implied that HnrnpK

binding does not require A-repeat, which we confirmed by ChIRP-western (Figure S6A). Thus, two sets of silencing proteins bind to different domains of Xist.

We reasoned that one or more of the A-repeat binding factors may be required for XCI. RNAi depletion in ES cells harboring wild-type Xist of each of these proteins, as well as Rnf40, a functional partner of Rnf20, showed that only depletion of Spn, but



**Figure 7. Xist A-Repeat Binds Spen, a Silencing Factor that Contributes to XCI**

(A) Similar proteins are enriched by ChIRP-MS of full-length Xist and A-repeat mutant, except three highlighted proteins: Wtap, Rnf20, and Spen. (B) siRNA depletion of the indicated factors show that only Spen is required for X-linked silencing of *Pgk1*.

(legend continued on next page)

not the other proteins, dramatically reduced Xist-mediated silencing of the X-linked gene *Pgk1* (Figure 7B). Rnf20 and Rnf40 depletion actually slightly increased *Pgk1* silencing, which is consistent with their known roles in enhancing transcription (Figure 7B) (Zhu et al., 2005). In addition, by two-color RNA-FISH, we found that Spen depletion results in reduced silencing of the X-linked genes *Mecp2* and *Rnf12*, with more frequent detection of nascent transcription from the Xist-coated inactive X chromosome in Spen-depleted cells compared to control cells or to cells depleted for Rnf20, Rnf40, or Wtap (Figures 7C and 7D). This experiment further illustrates that Spen is not apparently required for Xist RNA accumulation or spreading across the Xi but is specifically needed for transcriptional silencing, which is consistent with its specific association with the A-repeat region of Xist. Furthermore, we validated the requirement of Spen for silencing of *Grb10* in ESCs, where an Xist transgene is ectopically expressed on chr11 (Figure S6B). Collectively, these results suggest that Spen could be a functional mediator of Xist-RNA-driven gene silencing.

Spen is the mouse homolog of *Drosophila* homeotic mutant *Split ends* and encodes a transcriptional repressor (Arieti et al., 2014; Shi et al., 2001). Spen contains at least three RNA recognition motifs (RRMs) that can bind the lncRNA SRA to mediate RNA-directed transcriptional regulation (Arieti et al., 2014; Shi et al., 2001). Several existing Spen antibodies tested were not suitable for UV CLIP. Instead, we generated recombinant Spen RRM domains by in vitro translation and found that two or three of the Spen RRMs preferentially retrieved with Xist A-repeat over GFP mRNA in vitro (Figure S6C). These results suggest that Spen RRM domains may interact directly with the Xist A-repeat region.

## DISCUSSION

### ChIRP-MS: An RNA-Centric Interactome Technology

ChIRP-MS provides a potentially universal interactome discovery strategy that can be readily applied to any RNA of interest. We found comparable results from RNase-treated samples or isogenic cells that lack the target RNA, suggesting application in non-genetic systems. The use of different cross-linking reagents allows the investigator to potentially tune the degree of interactions captured from the target RNA. The thorough re-discovery of the spliceosome complex proteins by ChIRP-MS of U1 and U2 snRNPs validates the robustness of ChIRP-MS. In addition, the novel factors found in U1, U2, and Xist RBPs (functionally validated in the latter) demonstrate the added sensitivity of ChIRP-MS over traditional methods of RBP identification and provide a rich resource for future investigations. For example, U1-specific interaction with the cleavage stimulation and polyadenylation proteins has direct implications for “telescripting,” a critical process of U1-mediated protection from premature mRNA shortening (Berg et al., 2012; Kaida et al., 2010).

### Dynamic Plug-and-Play of Xist Binding Proteins

Our analysis revealed two sets of proteins that interact with Xist in a developmentally regulated manner. As Xist expression and reversibility of Xist-mediated gene silencing are tightly coupled to ESC differentiation, Xist may gain new silencing functions, perhaps through newly acquired or strengthened protein interactions, upon exit from pluripotency. Consistent with this idea, “Set 2” proteins bind Xist exclusively in differentiating ESCs (and EpiSCs and TSCs); this developmentally controlled assembly of Xist RNP provides a fail-safe backup for premature Xist expression during pluripotency. The expression of most factors in Set 2 remains stable throughout the differentiation of mESC into mEpiSC (< 2-fold change), as measured by whole-nucleus proteomic analysis (Song et al., 2012) (Figure S6D). Thus, the vast majority of Set 2 interactions are most parsimoniously explained by a change in Xist RNA that now allows interaction with a pre-existing set of proteins. In contrast, the compositions of Xist-RBPs are strikingly similar in differentiating ES cells, EpiSCs, and TSCs. TSCs are derived from extra-embryonic trophoblast cells, where the inactive X is always paternal (Takagi and Sasaki, 1975). It remains a standing debate in the field whether imprinted XCI differs from random XCI merely by a simple choice mechanism while sharing the same silencing machinery or whether the imprinted versus random XCI are fundamentally different. Our observations support the former hypothesis and suggest that the difference between random versus imprinted XCI is focused on the choice mechanism of the future Xi.

HnrnpU and HnrnpK emerged as the most enriched Xist-associated factors, and both functionally contribute to XCI. Although HnrnpU is required for Xist spreading across X chromosome, HnrnpK knockdown affects Xist-directed deposition of silencing histone modifications H2AK119ub and H3K27me3, the products of PRC1 and PRC2 complexes, respectively. Xist appears to directly bind PRC1, but not PRC2. This is consistent with recent reports demonstrating the PRC1-dependent recruitment of PRC2 complex (Blackledge et al., 2014; Cooper et al., 2014; Kalb et al., 2014). It has been reported that PRC2 binds specifically to the repeat A (repA) transcript of Xist, which is produced as a separate and shorter RNA (1.6 kb, including the A-repeat region) (Zhao et al., 2008), although the exact function of this shorter transcript remains unclear. One explanation for our findings could be the existence of different RNA isoforms with different functions. Further tests will be required to dissect the events by which Polycomb proteins associate with Xi.

### Modular Xist RNA Domains Link Spen- and HnrnpK-Mediated Silencing

Although the A-repeat was proposed to recruit PRC2 complex (Zhao et al., 2008), PRC2 itself is dispensable for the initiation of gene silencing during XCI (Kalanitry and Magnuson, 2006). Furthermore, in the A-repeat deletion Xist mutant, PRC2 and

(C) siRNA depletion of Spen interferes with XCI in cells, as indicated by co-localization of Xist “cloud” and active transcription of X-linked genes *Rnf12* and *Mecp2* (arrowheads) on the same chromosome.

(D) Quantification of cells with expression of *Mecp2* and *Rnf12* on the Xist-coated chromosome by counting >100 cells from 3 replicates. A proportion of cells does not upregulate Xist and does not coat (around 40%); we counted only the cells with Xist domains.

(E) Model of the cell-state- and scaffold-specific loading of Xist-RBPs and their chromatin-modifying functions.



H3K27me3 are still recruited to the Xist-coated chromosome (da Rocha et al., 2014; Plath et al., 2003). Imaging studies suggested that Xist RNA create a transcriptionally inactive nuclear compartment, independent of the A-repeat, but that the A-repeat is required for the movement of genes into this compartment as they become silenced (Chaumeil et al., 2006). These observations suggest that factors beyond PRC2 are at play.

Our results revealed the A-repeat—essential for Xist-mediated gene silencing (Wutz et al., 2002)—as a key element for the developmentally regulated binding of several proteins. The selective abrogation of three protein interactions but full preservation of all others by the A-repeat deletion highlights the modular organization of Xist. We found Spen, a potent transcriptional repressor, to be important for Xist-mediated silencing. Spen interaction with Xist is increased upon differentiation, suggesting a gain of Spen-associated silencing activity to the Xist RNP. The *Spen* knockout is embryonic lethal at E12.5 (Kuroda et al., 2003), which is later than expected if XCI is fully defective. However, the knockout was not performed with a maternal germline depletion of the protein, so an earlier phenotype masked by the maternal pool cannot be ruled out. On the other hand, Spen may well collaborate with other Xist-recruited silencing activities, and there may also be potential redundancy with two other mammalian *Spit ends* homologs.

The reported association between Spen and MBD3-NuRD complex nominates several gene-silencing pathways, including ATP-dependent nucleosome remodeling, histone deacetylation via HDACs, and modulation of DNA methylation (Shi et al., 2001; Zhang et al., 1999). NuRD complex decommissions ESC enhancers to enable differentiation and lineage commitment—the same developmental window where XCI takes place (Reynolds et al., 2012; Whyte et al., 2012). It is conceptually appealing that the same silencing mechanism that turns off pluripotency regulators may both enable Xist expression (by removing repression of Xist) and endow Xist with the silencing power to achieve XCI. Intriguingly, Spen interacts with Mbd3 (Shi et al., 2001); NuRD recruitment to active enhancers is believed to occur through Mbd3 recognition of 5-hydroxymethylcytosine (Yildirim et al., 2011). NuRD-mediated deacetylation of H3K27ac also permits PRC2-mediated H3K27me3 and gene silencing (Reynolds et al., 2012). Thus, the combination of NuRD and Polycomb activity can turn an active gene into an inactive one. We propose that Xist may serve as a physical scaffold for organizing at least two chromatin modification activities—a writer to deposit silencing marks via PRC1 and an eraser to remove active marks via Spen and associated factors—that, together, coordinately enforce permanent epigenetic silencing (Figure 7E).

Although the other two A-repeat associating factors do not directly impact XCI in our limited analysis, they could conceptually still contribute to XCI. Rnf20 is the E3 ubiquitin ligase for H2BK120ub1, a histone modification that marks the gene bodies of transcriptionally active genes (Zhu et al., 2005). Xist has been proposed to preferentially target actively transcribed genes on X chromosome, exploiting the spatial proximity of actively transcribed loci to efficiently target Xist-associated silencing factors (Engreitz et al., 2013; Simon et al., 2013). Furthermore, the A-repeat mutant of Xist shows reduced binding to such active regions (Engreitz et al., 2013), which may be explained by the

inability of the Xist A-repeat mutant to seek out Rnf20 complex loaded on active loci. Finally, Wtap is involved in the installation of the N6-methyladenosin (m6A) on RNAs. Wtap binding to the A-repeat of Xist is consistent with the presence of m6A in the same region of the RNA (data not shown). The functional impact of Wtap binding or m6A modification remains to be understood but represents an exciting perspective given the strategic importance of the domain in question. Our results set the stage for future structure-function analysis of Xist and its interacting proteins as a paradigm to understand functional motifs in lncRNAs.

## EXPERIMENTAL PROCEDURES

### ChIRP-MS

10–20 15 cm dishes of cells were used per ChIRP-MS experiment (100 million–500 million cells, depending on the cell type). Cell harvesting, lysis, disruption, and ChIRP were essentially performed as previously described (Chu et al., 2012), with the following modifications: (1) cells are cross-linked in 3% formaldehyde for 30 min, followed by 0.125 M glycine quenching for 5 min; (2) hybridization can be started late in the day and left running overnight to reduce hands-on time; (3) for MS experiments, lysates were pre-cleared by incubating with 30  $\mu$ l washed beads per ml of lysate at 37°C for 30 min with shaking (prior to hybridization, beads were removed twice from lysate using a magnetic stand); (4) for RNase control, lysates are pooled first and aliquoted into two equal amounts. 1/1,000 volume of 10 mg/ml RNase A (Sigma) is added to the RNase control sample, and both control and non-treated samples are incubated at 37°C for 30 min with mixing prior to hybridization steps. This can be done concurrently with pre-clearing. RNA extraction can be performed from a small aliquot of post-ChIRP beads as described (Chu et al., 2012). For protein elution, beads were collected on magnetic stand, resuspended in biotin elution buffer (12.5 mM biotin [Invitrogen], 7.5 mM HEPES [pH 7.5], 75 mM NaCl, 1.5 mM EDTA, 0.15% SDS, 0.075% sarkosyl, and 0.02% Na-Deoxycholate), mixed at room temperature (r.t.) for 20 min and at 65°C for 10 min. Eluent was transferred to a fresh tube, and beads were eluted again. The two eluents were pooled, and residual beads were removed again using the magnetic stand. 25% total volume TCA was added to the clean eluent, and after thorough mixing, proteins were precipitated at 4°C overnight. The next day, proteins were pelleted at 16,000 rcf at 4°C for 30 min. Supernatant was carefully removed from the belly side of tubes, and protein pellets on the spine of tubes (sometimes invisible at this step) were washed once with cold acetone and pelleted again at 16,000 rcf at 4°C for 5 min, and acetone was removed. Pellets (much more visible now) were briefly centrifuged again and, after removal of residual acetone, were left to air-dry for 1 min on bench-top. Proteins are then immediately solubilized in desired volumes of 1× Laemmli sample buffer (Invitrogen) and boiled at 95°C for 30 min with occasional mixing for reverse-crosslinking. Final protein samples were size-separated in bis-tris SDS-PAGE gels (Invitrogen) for western blots or MS. See [Extended Experimental Procedures](#) and [Table S2](#) for ChIRP probe design.

### Defining Proteins Identified by ChIRP-MS

Potential MS artifacts were first filtered by removing low-confidence protein hits with fewer than 9 peptides from a single gel-C slice and fewer than 16 total peptides (a simpler cut-off of >10 peptides from any single gel-C slice was used for U1/U2). Thereafter, a stringent cut-off of  $\log_2 \geq 3.5$  between experiment and control ( $\geq 11.3$  fold enrichment) is applied to eliminate RNA-independent background interactions. Specific hits of 7SK ChIRP-MS will be reported elsewhere. To define specific versus non-specific components of the Xist lncRNP, ChIRP-MS hits from Xist (differentiated ESC), U1, U2, and 7SK were first ranked based on peptide abundance. Xist-specific interactors are defined as proteins with Xist ChIRP-MS rank at least twice better than in ChIRP-MS of U1, U2, and 7SK. Non-specific interactors are proteins that show rank ratio < 2 in Xist ChIRP versus U1, U2, or 7SK. For the purpose of comparison, mouse protein names of 7SK and Xist hits were replaced with their human counterparts (no ambiguity).

### Defining Xist-Specific RBPs versus Promiscuous RBPs

The most enriched protein (most peptide counts in experiment) is ranked 1, the second most enriched is ranked 2, and so forth. “Specific interactors” for Xist are defined as proteins that have a rank that is at least 2-fold better than in all three other ChIRP-MS of U1, U2, or 7SK.

### Knockdown Studies

siRNAs and shRNAs are purchased from Dharmacon and Invitrogen. Transfection was performed with nucleofector or RNAiMAX. See [Extended Experimental Procedures](#) and [Table S5](#) for full details.

### Microscopy

Xist-FISH, Usp9x-FISH, and co-IF are performed with sm-FISH probes with standard protocol. All other dual-color FISH were essentially performed as previously described ([Chaumeil et al., 2008](#)). See [Extended Experimental Procedures](#) for full protocols and the list of reagents used.

### RNA Crosslinking IP and Interaction Studies

Clip-qRT-PCR was essentially performed as described ([Flynn et al., 2015](#)) and triple flag-tagged codon-optimized 2× RRM and 3× RRM Spen fragments were used in vitro interaction studies. See [Extended Experimental Procedures](#) for full details.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.025>.

### ACKNOWLEDGMENTS

We thank A. Wutz (ETH), A. Surani, J. Zyllicz (University of Cambridge), A. Smith (University of Cambridge), and A. Gillich (Stanford) for reagents and advice; W. Lane and R. Robinson (Harvard) for MS analysis; A. Olson for assistance with confocal microscopy (Stanford Neuroscience Microscopy Service supported by NIH NS069375); and P. Walker (Stanford PAN facility) for ChIRP oligo synthesis. This work was supported by NIH P50-HG007735 and California Institute for Regenerative Medicine. H.Y.C. is an Early Career Scientist of the Howard Hughes Medical Institute. E.H. is supported as an Equipe labellisée “La Ligue Contre Le Cancer” (Equipe Labellisé to E.H.) and by Labex DEEP (ANR-11-LBX-0044) part of the IDEX IDEX PSL (ANR-10-IDEX-000102 PSL), as well as the EpiGeneSys FP7 257082 Network of Excellence and ERC Advanced Investigator award 250367. H.Y.C. is a founder of Epinomics, Inc. and is a member of the Scientific Advisory Board of RaNA Therapeutics.

Received: October 12, 2014

Revised: January 29, 2015

Accepted: March 11, 2015

Published: April 2, 2015

### REFERENCES

- Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499, 360–363.
- Arieti, F., Gabus, C., Tambalo, M., Huet, T., Round, A., and Thore, S. (2014). The crystal structure of the Split End protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs. *Nucleic Acids Res.* 42, 6742–6752.
- Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., and Dreyfuss, G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. *Cell* 150, 53–64.
- Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* 157, 1445–1459.
- Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature* 368, 154–156.
- Buzin, C.H., Mann, J.R., and Singer-Sam, J. (1994). Quantitative RT-PCR assays show Xist RNA levels are low in mouse female adult tissue, embryos and embryoid bodies. *Development* 120, 3529–3536.
- Calabrese, J.M., Sun, W., Song, L., Mugford, J.W., Williams, L., Yee, D., Starmer, J., Mieczkowski, P., Crawford, G.E., and Magnuson, T. (2012). Site-specific silencing of regulatory elements as a mechanism of X inactivation. *Cell* 151, 951–963.
- Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* 20, 2223–2237.
- Chaumeil, J., Augui, S., Chow, J.C., and Heard, E. (2008). Combined immunofluorescence, RNA fluorescent in situ hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation. *Methods Mol. Biol.* 463, 297–308.
- Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell* 44, 667–678.
- Chu, C., Quinn, J., and Chang, H.Y. (2012). Chromatin isolation by RNA purification (ChIRP). *J. Vis. Exp.* 61, 3912.
- Chu, C., Spitale, R.C., and Chang, H.Y. (2015). Technologies to probe functions and mechanisms of long noncoding RNAs. *Nat. Struct. Mol. Biol.* 22, 29–35.
- Clemson, C.M., McNeil, J.A., Willard, H.F., and Lawrence, J.B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.* 132, 259–275.
- Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep.* 7, 1456–1470.
- Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R., and Jaenisch, R. (1999). Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. *Nat. Genet.* 22, 323–324.
- da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 Is Implicated in the Initial Xist-Induced Targeting of PRC2 to the Inactive X Chromosome. *Mol. Cell* 53, 301–316.
- de Napolés, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* 7, 663–676.
- Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* 341, 1237973.
- Fang, J., Chen, T., Chadwick, B., Li, E., and Zhang, Y. (2004). Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. *J. Biol. Chem.* 279, 52812–52815.
- Flynn, R.A., Martin, L., Spitale, R.C., Do, B.T., Sagan, S.M., Zarnegar, B., Qu, K., Khavari, P.A., Quake, S.R., Sarnow, P., and Chang, H.Y. (2015). Dissecting noncoding and pathogen RNA-protein interactomes. *RNA* 21, 135–143.
- Gendrel, A.V., and Heard, E. (2011). Fifty years of X-inactivation research. *Development* 138, 5049–5055.
- Gesteland, R.F., and Atkins, J.F. (1993). *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World* (Cold Spring Harbor Laboratory Press).
- Gillich, A., Bao, S., Grabole, N., Hayashi, K., Trotter, M.W., Pasque, V., Magúsdóttir, E., and Surani, M.A. (2012). Epiblast stem cell-based system reveals reprogramming synergy of germline factors. *Cell Stem Cell* 10, 425–439.

- Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A. (2009). Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* 136, 1063–1069.
- Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., Tsutui, K., and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev. Cell* 19, 469–476.
- Hayashi, K., and Surani, M.A. (2009). Resetting the epigenome beyond pluripotency in the germline. *Cell Stem Cell* 4, 493–498.
- Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468, 664–668.
- Kalanitry, S., and Magnuson, T. (2006). The Polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet.* 2, e66.
- Kalb, R., Latwiel, S., Baymaz, H.I., Jansen, P.W., Müller, C.W., Vermeulen, M., and Müller, J. (2014). Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* 21, 569–571.
- Keohane, A.M., O'Neill, L.P., Belyaev, N.D., Lavender, J.S., and Turner, B.M. (1996). X-Inactivation and histone H4 acetylation in embryonic stem cells. *Dev. Biol.* 180, 618–630.
- Kuroda, K., Han, H., Tani, S., Tanigaki, K., Tun, T., Furukawa, T., Taniguchi, Y., Kurooka, H., Hamada, Y., Toyokuni, S., and Honjo, T. (2003). Regulation of marginal zone B cell development by MINT, a suppressor of Notch/RBP-J signaling pathway. *Immunity* 18, 301–312.
- Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annu. Rev. Genet.* 42, 733–772.
- Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300, 131–135.
- Pomeranz Krummel, D.A., Oubridge, C., Leung, A.K., Li, J., and Nagai, K. (2009). Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* 458, 475–480.
- Rastan, S. (1982). Timing of X-chromosome inactivation in postimplantation mouse embryos. *J. Embryol. Exp. Morphol.* 71, 11–24.
- Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., Mosaku, O., Signolet, J., Brennecke, P., Kalkan, T., et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* 10, 583–594.
- Riley, K.J., and Steitz, J.A. (2013). The "Observer Effect" in genome-wide surveys of protein-RNA interactions. *Mol. Cell* 49, 601–604.
- Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166.
- Ruepp, A., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Stransky, M., Waegle, B., Schmidt, T., Doudieu, O.N., Stümpflen, V., and Mewes, H.W. (2008). CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* 36, D646–D650.
- Shi, Y., Downes, M., Xie, W., Kao, H.Y., Ordentlich, P., Tsai, C.C., Hon, M., and Evans, R.M. (2001). Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev.* 15, 1140–1151.
- Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* 504, 465–469.
- Song, J., Saha, S., Gokulrangan, G., Tesar, P.J., and Ewing, R.M. (2012). DNA and chromatin modification networks distinguish stem cell pluripotent ground states. *Mol. Cell. Proteomics* 11, 1036–1047.
- Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 409, 539–542.
- Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature* 256, 640–642.
- Takagi, N., Sugawara, O., and Sasaki, M. (1982). Regional and temporal changes in the pattern of X-chromosome replication during the early post-implantation development of the female mouse. *Chromosoma* 85, 275–286.
- Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* 148, 664–678.
- West, J.A., Davis, C.P., Sunwoo, H., Simon, M.D., Sadreyev, R.I., Wang, P.I., Tolstorukov, M.Y., and Kingston, R.E. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol. Cell* 55, 791–802.
- Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225.
- Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* 5, 695–705.
- Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* 30, 167–174.
- Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147, 1498–1510.
- Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* 13, 1924–1935.
- Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556–560.
- Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* 419, 182–185.
- Zhu, B., Zheng, Y., Pham, A.D., Mandal, S.S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005). Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Mol. Cell* 20, 601–611.

# Retraction Notice to: A Pleiotropically Acting MicroRNA, miR-31, Inhibits Breast Cancer Metastasis

Scott Valastyan, Ferenc Reinhardt, Nathan Benaich, Diana Calogrias, Attila M. Szász, Zhigang C. Wang, Jane E. Brock, Andrea L. Richardson, and Robert A. Weinberg\*

\*Correspondence: [weinberg@wi.mit.edu](mailto:weinberg@wi.mit.edu)  
<http://dx.doi.org/10.1016/j.cell.2015.03.047>

(Cell 137, 1032–1046; June 11, 2009)

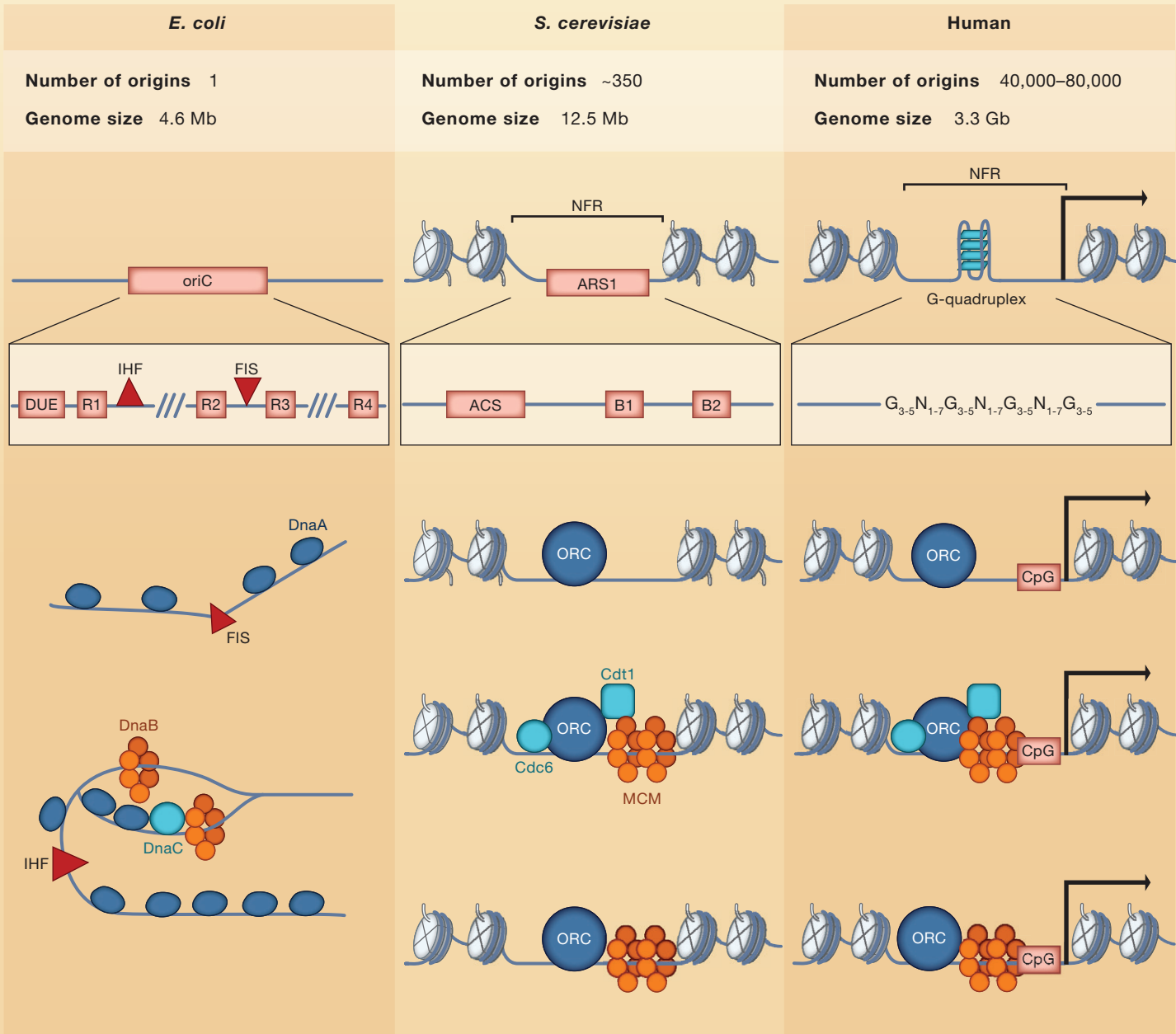
Our study reported that miR-31 is a regulator of multiple mRNAs important for different aspects of breast cancer metastasis. We recently identified concerns with several figure panels in which original data were compiled from different replicate experiments in order to assemble the presented figure. The scope of the figure preparation issues includes compiling data from independent experiments to present them as one internally controlled experiment, statistical analyses based on technical replicates that are not reflective of the biological replicates, and comparisons of selectively chosen data points from multiple experiments. As many of the published figures are therefore not appropriate or accurate representations of the original data, we believe that the responsible course of action is to retract the paper. We apologize for any inconvenience we have caused.



# SnapShot: Origins of DNA Replication

Rachel L. Creager, Yulong Li, and David M. MacAlpine  
Department of Pharmacology and Cancer Biology, Duke University Medical Center,  
Durham, NC 27710, USA

Cell



Histone PTM	Enzyme	Function	Organism
H4K20me1	PR-Set7	Promotes pre-RC assembly	Human, mouse
H4K20me2	Suv4-20h1/2	ORC recruitment via ORC1 BAH domain	Metazoa
H3K4me2	COMPASS complex	Origin activation	Yeast
H3K27me	ATXR5, ATRX6	Represses re-replication of heterochromatic origins	Arabidopsis
H3K36me	Set2	Regulates Cdc45 association with origins	Yeast
H3K79me2	DOT1L	Enriched at origins, loss of DOT1L leads to re-replication	Human
H3K4me3 demethylation	KDM5C/JARID1C	Promotes early-origin activation	Human
H4Ac	Hbo1	Promotes pre-RC assembly	Human, xenopus, fly
Bulk H3 and H4 acetylation	Multiple HATs	Regulates developmentally programmed origin of the $\beta$ -globin locus	Human
H4K5Ac/H4K12Ac	Hat1p/Hat2p	Hat1/Hat2 interact with and enhance the function of ORC	Yeast
H4K16Ac	MOF	Promotes male-specific early-origin activation on the X chromosome	Fly
H3,H4 deacetylation	Rpd3	Delays late-origin activation, developmental transition in origin specificity	Yeast, fly
H3,H4 deacetylation	Sir2p	Inhibits pre-RC assembly at a subset of origins	Yeast
H4K5Ac deacetylation	Sum1/Rfm1/Hst1	Enhances efficiency of replication initiation at a subset of origins	Yeast
H2B Ub	Bre1	Enriched at origins, impacts fork elongation	Yeast

# SnapShot: Origins of DNA Replication

Cell

Rachel L. Creager, Yulong Li, and David M. MacAlpine

Department of Pharmacology and Cancer Biology, Duke University Medical Center,  
Durham, NC 27710, USA

More than 50 years ago, Jacob et al. (1964) proposed an elegant model for the regulation of DNA replication in bacteria. In the replicon model, the fundamental unit of DNA replication, the replicon, would be governed by a *cis*-acting replicator sequence and a *trans*-activating initiator factor. Despite the increased size and complexity of eukaryotic genomes, eukaryotic DNA replication continues to be guided by the fundamental principles and concepts established in the replicon model. Eukaryotic origins of replication (replicators) are defined by *cis*-acting sequences or structural DNA elements that are recognized by functionally conserved *trans*-acting initiator factors (ORC, Cdc6, Cdt1, and Mcm2-7).

## Prokaryotic Replication Origins

*E. coli* has a single circular chromosome that is ~4.6 megabases in length, containing a single origin of replication (*oriC*). The two major *cis*-acting features of the ~250 bp *oriC* are an AT-rich DNA unwinding element (DUE) and multiple 9 bp DnaA-binding motifs (Skarstad and Katayama, 2013).

DnaA is a AAA+ ATPase that recognizes both high- and low-affinity binding sites throughout *oriC*. High-affinity DnaA binding sites (R1–R4) are occupied throughout most of the cell cycle; however, low-affinity binding sites become occupied only at replication initiation. Two accessory proteins, Fis and IHF, repress or stimulate initiation of DNA replication, respectively, by altering the conformation of DNA surrounding *oriC*. Binding of a full complement of DnaA molecules (10–20) leads to DNA unwinding at DUE. This unwinding stimulates DnaC, a AAA+ ATPase helicase loader, to assemble two homohexamers of the DnaB helicase complex on the newly melted DNA, forming the pre-replicative complex (pre-RC) (Costa et al., 2013).

## Eukaryotic DNA Replication Origins

A consequence of the large size of eukaryotic genomes is that multiple origins of replication must be utilized in parallel to facilitate the complete duplication of the genome within the confines of S phase (Gilbert, 2004). There are ~350 origins of replication distributed throughout the *S. cerevisiae* genome. In contrast, there are an estimated 40,000–80,000 origins distributed throughout the much larger human genome.

As in bacteria, both *cis*- and *trans*-acting factors define start sites of eukaryotic DNA replication. Eukaryotic *cis*-acting replicator elements were first identified in the model organism *S. cerevisiae* as autonomous replicating sequence (ARS) elements of ~200 bp. Each ARS element contains a conserved ARS consensus sequence (ACS), which, at its core, is an 11 bp T-rich motif that is necessary but not sufficient for origin function. In addition to the ACS, there are several other poorly defined B sequence elements that contribute to helicase loading and DNA unwinding. In contrast, conserved replicator sequences that direct origin selection in higher eukaryotes have remained elusive. Recently, low-complexity GC-rich sequences that are able to form G-quadruplexes have been identified as potential replicator elements, suggesting that DNA secondary structure may play a role in origin licensing in higher eukaryotes (Cayrou et al., 2012).

There is remarkable functional conservation between prokaryotic and eukaryotic *trans*-acting initiator factors. Analogous to DnaA, the origin recognition complex (ORC) recognizes and binds to replicator origin sequences throughout the majority of the cell cycle. ORC is composed of six subunits, five of which are AAA+ ATPases. In G1 of the cell cycle, another initiation factor and AAA+ ATPase, Cdc6, associates with ORC and coordinates the loading of a double hexamer of the minichromosome maintenance (Mcm2–7) helicase complex via an interaction with Cdt1 to form the pre-RC (Bell and Kaguni, 2013). Formation of the pre-RC “licenses” the origin for potential activation in the subsequent S phase.

Eukaryotic *cis*-acting replicator elements are necessary but not sufficient for origin activity, as there are many more ACS motif matches and potential G-quadruplexes than utilized origins of replication. Epigenetic features, including promoters, CpG islands, nucleosome organization, and post-translational modification of histones, also impact the selection and activation of eukaryotic replication origins (Ding and MacAlpine, 2011). A nucleosome-free region (NFR) with well-positioned flanking nucleosomes is a conserved feature of eukaryotic replicator sequences, and maintenance of the nucleosome free region at the origin is critical for function. Eukaryotic origins are typically associated with intergenic sequences. In higher eukaryotes, origins and ORC binding sites are frequently found in the NFR associated with transcription start sites and CpG islands, whereas in *S. cerevisiae*, ORC is excluded from transcription start sites.

A diverse array of histone PTMs are correlated with ORC binding and origin function; however, only a handful of epigenetic marks have been mechanistically linked to specific steps in origin selection and activation (Dorn and Cook, 2011; Méchali et al., 2013). The methylation state of histone H4 lysine 20 in metazoans has been linked to ORC binding and pre-RC assembly. Dimethylation of H4K20 by the methyltransferase Suv4-20h1/2 is recognized by the bromo-adjacent homology (BAH) domain of ORC1. In addition, monomethylation of H4K20 by the cell-cycle-regulated methyltransferase, PR-Set7, promotes pre-RC assembly. The histone acetyltransferase (HAT) Hbo1 interacts with ORC, Cdt1, and Mcm2 and is required for efficient pre-RC assembly. Presumably, Hbo1 acetylates origin-proximal nucleosomes on histone H4 to promote pre-RC assembly; however, it remains possible that the target of Hbo1 is not histones but instead specific pre-RC components. Finally, although numerous epigenetic modifications and chromatin states have been correlated with origin function, it is important to stress that correlation does not equal causation. The *S. cerevisiae* histone deacetylase Rpd3 represses global origin activation not by deacetylating histone H3 in the vicinity of replication origins but rather by regulating silencing of the origin-rich multicopy rDNA locus, which serves as a sink for sequestering key replication initiation factors (Yoshida et al., 2014).

## ACKNOWLEDGMENTS

We thank members of the MacAlpine group for ideas and suggestions. This work was supported by NIH grants R01 GM104097 (D.M.M.) and T-32 CA059365 (R.L.C.).

## REFERENCES

- Bell, S.P., and Kaguni, J.M. (2013). Cold Spring Harb. Perspect. Biol. 5, a010124.
- Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., and Méchali, M. (2012). Cell Cycle 11, 658–667.
- Costa, A., Hood, I.V., and Berger, J.M. (2013). Annu. Rev. Biochem. 82, 25–54.
- Ding, Q., and MacAlpine, D.M. (2011). Crit. Rev. Biochem. Mol. Biol. 46, 165–179.
- Dorn, E.S., and Cook, J.G. (2011). Epigenetics 6, 552–559.
- Gilbert, D.M. (2004). Nat. Rev. Mol. Cell Biol. 5, 848–855.
- Jacob, F., Brenner, S., and Cuzin, F. (1964). Cold Spring Harb. Symp. Quant. Biol. 28, 329.
- Méchali, M., Yoshida, K., Coulombe, P., and Pasero, P. (2013). Curr. Opin. Genet. Dev. 23, 124–131.
- Skarstad, K., and Katayama, T. (2013). Cold Spring Harb. Perspect. Biol. 5, a012922.
- Yoshida, K., Bacal, J., Desmarais, D., Padiou, I., Tsaponina, O., Chabes, A., Pantescio, V., Dubois, E., Parrinello, H., Skrzypczak, M., et al. (2014). Mol. Cell 54, 691–697.